Cross-Gender Social Ties Around the World

By Michael Bailey, Drew Johnston, Theresa Kuchler, Ayush Kumar, Johannes Stroebel*

Economists, sociologists, and other social scientists have long been interested in the determinants and effects of cross-gender social ties (e.g., McPherson, Smith-Lovin and Cook, 2001; Currarini, Jackson and Pin, 2009). However, due to a lack of large-scale representative data on social networks, empirical work has usually studied social ties in a single setting or geography, precluding a systematic analysis of any spatial or other variation in cross-gender social connections.

Here, we introduce, analyze, and publicly release a new global dataset on cross-gender friendship links at the sub-national level for nearly 200 countries and territories. Our measures are based on more than 1.38 trillion social ties observed between over 1.8 billion users on Facebook, a global online social networking service.

The aggregated data is available for download at the HDX (`https://data.humdata.org/dataset/cross-gender-ties`). We hope that it will facilitate new research to understand the dynamics that shape the formation of cross-gender social ties, as well as the effects of such ties.

I. Data and Empirical Strategy

We study Facebook users between 18 and 65 years old who were active on the platform in the 30 days before January 9, 2025. We observe their predicted home location,[1] their self-reported gender, and their links to other users, which Facebook refers to as 'friendships.' Prior work has shown that these Facebook friendships generally reflect real-world social ties (Bailey et al., 2018, 2020, 2021, 2024; Chetty et al., 2022*a*,*b*). To focus on active users, we drop users in the bottom 25% of each country's friend count distribution.

For each friendship, we observe a measure of tie strength based on a proprietary model developed by Facebook, which we use to rank each individual's friendships from strongest to weakest.

To construct measures of cross-gender social ties by location, we first define FS_i^n as the share of individual i's top-n same-country friends who are female. We then define the Cross-Gender Friending Ratio (CGFR) in location c as the ratio of female friends in men's networks to the share of female friends in women's networks.

$$\text{CGFR}_c^n = \frac{\sum_{i \in m \cap c} \text{FS}_i^n / |m \cap c|}{\sum_{i \in f \cap c} \text{FS}_i^n / |f \cap c|} \quad (1)$$

This normalization controls for spatial variation in the share of Facebook users from each gender as well as differences in the average number of Facebook friends by gender.

In places where men and women form equal shares of their ties with women, the value of the CGFR is 1. Values of the CGFR less than 1 indicate the presence of gender-segregated social networks, which could be driven by both preferences—specifically the degree of "gender homophily"—and institutional or legal environments.

While the CGFR has several properties that make it appealing as a globally-uniform sub-national measure of cross-gender social ties, it is important to note that it only captures social links among Facebook users, complicating any comparison of absolute numbers across regions with potentially different selection into using the platform. Despite such concerns, we show below that the CGFR correlates strongly with several external measures of cross-gender relations.

* Bailey: Meta Platforms (mcbailey@meta.com). Johnston: Meta Platforms (djohnston@meta.com). Kuchler: NYU Stern (tkuchler@stern.nyu.edu). Kumar: NYU Stern (ayush.kumar@stern.nyu.edu). Stroebel: NYU Stern (johannes.stroebel@nyu.edu). This research was facilitated through a research consulting agreement between the academic authors and Meta Platforms.

[1] Home locations are estimated using several signals, including on-platform activity, self-reported location, and the IP addresses used to connect to the platform.

II. Cross-Gender Friending

In Figure 1, we plot the $CGFR_c^n$ across four countries and values of n between 1 and 300. In each country, the $CGFR_c^n$ is below 1 for all values of n, indicating the existence of gender-segregation across various definitions of individuals' social networks. Gender segregation also varies by tie strength. An individuals' closest social ties are generally the least selected by gender, perhaps reflecting that these ties often correspond to family members, where gender homophily and institutional constraints to cross-gender ties may be less prevalent than in other contexts.

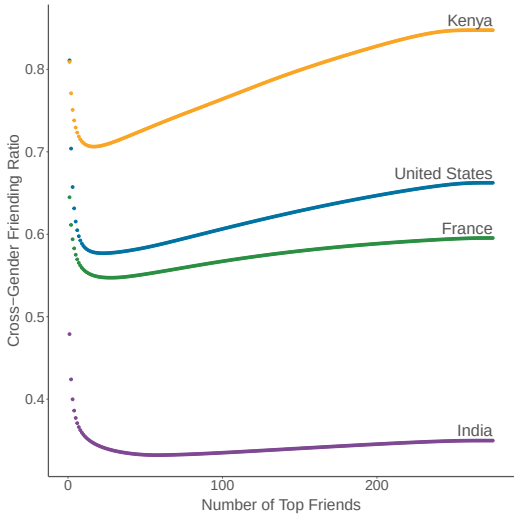

Figure 1. Cross-Gender Friending Ratio

Social ties generally display the strongest gender segregation within the top 25 or so friends. Networks become less gender-segregated when they are defined to include a broader set of friends, with the largest such changes in Kenya and the smallest in India. Figure 1 also highlights differences in the gender composition of networks across countries. In India, the average male has only 34.5% as many female friends among his 200 closest ties as the average woman does; in Kenya, this ratio is 82.8%.

Figure 2 shows the CGFR_c^{200} across subnational regions for multiple continents, using a common color scale across the various maps.[2] There is substantial variation in cross-gender connectedness both across and within countries: about 49% of the global cross-region variation is within country.

Rates of cross-gender friending are higher in Central America than in the U.S. and Canada. In Europe, Germany, France, and Czechia have relatively high levels of cross-gender friending, while Turkey and the Balkans (and to a lesser extent countries in Scandinavia) have lower rates. In Africa, North African counties have the lowest rates of cross-gender friendships; in South America, cross-gender friendship links are least common in Brazil and Argentina. In Southeast Asia, Myanmar has the highest and Vietnam the lowest rate of cross-gender friending. Cross-gender friendships are relatively rarer in much of South Asia, with the notable exceptions of Nepal and Bhutan.

Figure 3 shows that the observed cross-country variation in the CGFR correlates strongly ($\rho = -0.75$) with gender differences in labor force participation, as provided by the UN's Gender Inequality Index (UNDP, 2024), suggesting that both measures pick up related aspects of the broader environment driving cross-gender relationships. This finding raises our confidence that the CGFR captures real differences in gender relations rather than differences in Facebook usage behavior by gender.

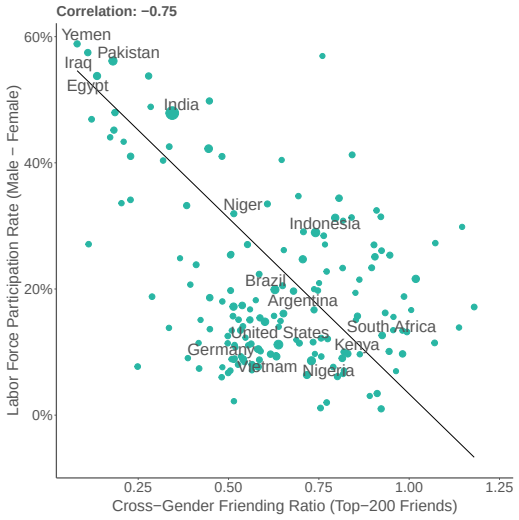

Figure 3. LFPR Gap vs. CGFR

[2]We define regions using definitions provided by version 4.1 of GADM. For most countries, we use the second-level granularity, which corresponds to counties in the U.S.; for the few countries where this granularity is unusually large or small, we use a different level.

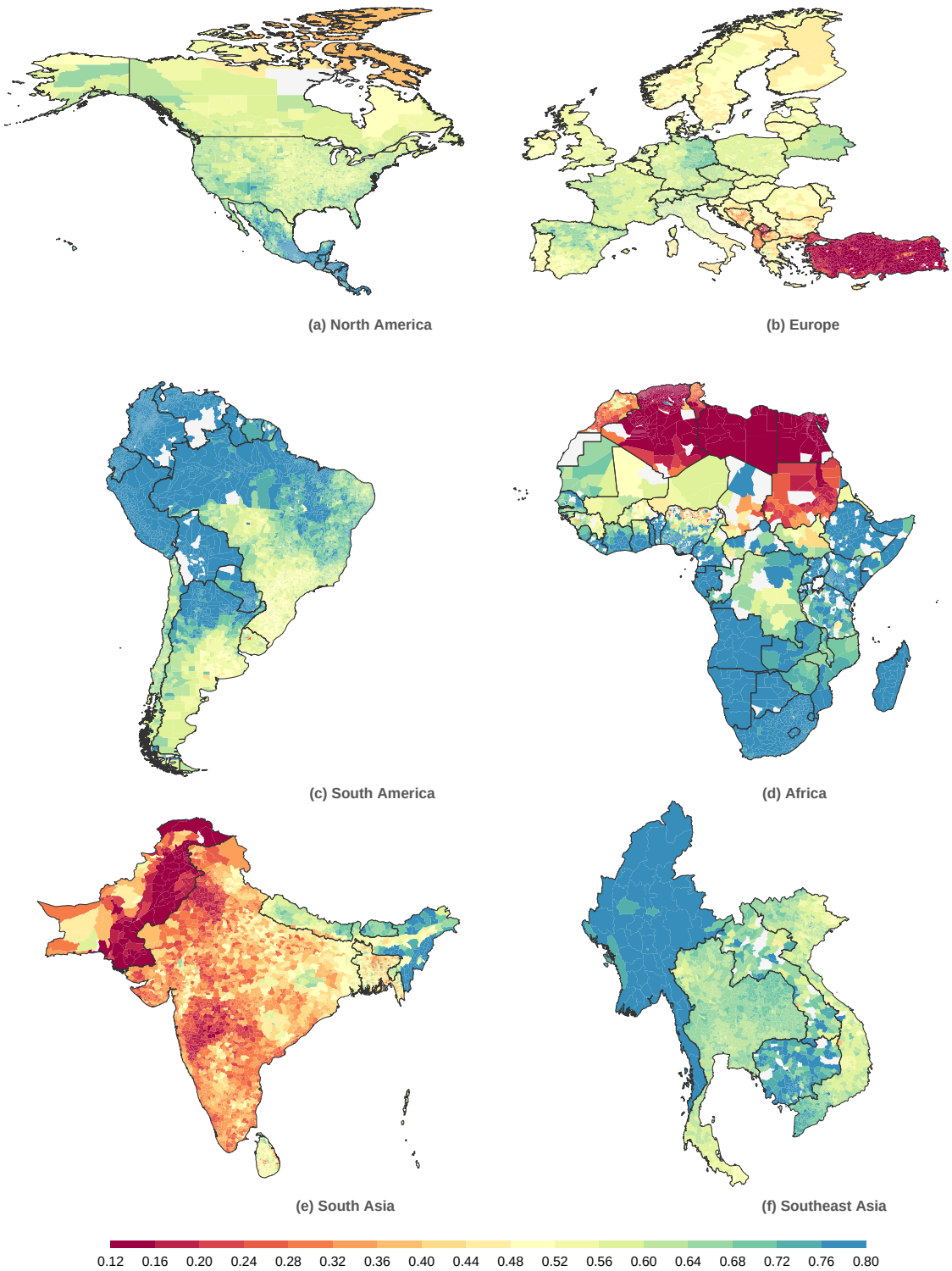

Figure 2. Maps of Cross-Gender Friending Ratio over Top-200 Friends

Note: Figure shows subnational variation in the CGFR across the world. Countries and regions with insufficient data are left blank. Maps of other areas of the world (Indonesia, Australia, Maritime Southeast Asia, and the Middle East), as well as maps colored by within-region deciles of the CGFR are available in the Online Appendix.

Table 1—Cross-Gender Friending Ratio and Beliefs on Gender Issues

	University more important for men (1)	Men have greater right to work (2)	Support women's rights (3)	Men make better political leaders (4)
CGFR (Top-200)	-0.133**	-0.170***	0.287***	-0.198***
	(0.061)	(0.049)	(0.066)	(0.050)
Demographic Group × Country FE	Yes	Yes	Yes	Yes
Observations	87,372	88,019	84,047	86,251
R^2	0.102	0.060	0.063	0.096

Note: Each column shows a regression of a measure of attitudes towards women from the World Values Survey (WVS) and the European Values Survey (EVS) against measures of CGFR in the locations of the respondents. All dependent variables are normalized to have mean 0 and standard deviation 1 within each country. CGFR is also normalized within-country, with each region weighted by its sample size. We use the 2016 NUTS3 regions from the restricted EVS dataset and GADM2 regions based on coordinates from the seventh wave of the WVS to assign CGFR values to survey respondents. Each model includes fixed effects for the country of the respondent, interacted with the following demographic categories observed in the surveys: religious group, subjective income decile, age group, marital status, immigration status, gender, employment status, and education level. The dependent variables correspond to different questions from the WVS. Each question asks respondents if they agree with a particular statement. The statements for questions are: (1) "University is more important for a boy than for a girl" (1-4, strongly disagree to strongly agree), (2) "When jobs are scarce, men should have more right to a job than women" (1-5, strongly disagree to strongly agree), (3) "Essential feature of democracy: women have the same rights as men" (0-10, against democracy to essential characteristic of democracy), (4) "On the whole, men make better political leaders than women do" (1-4, strongly disagree to strongly agree). Standard errors clustered at the GADM2/NUTS3 region level appear in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

One benefit of observing the networks of over 1.8 billion people is that we can also produce precise measures of sub-national variation in the gender composition of networks.

Figure 2 shows that within the United States, there is relatively little variation in cross-gender friending rates, though there is somewhat more gender segregation across the Great Plains and Appalachia. In Germany, there is less social gender segregation in the former East Germany than in the former West Germany. In both Spain and Italy, the north of the country has less gender segregation in networks than the south. In India there is substantial gender segregation in friendship networks in much of the country, though rates of cross-gender friending are higher in the eastern states of Nagaland, Mizoram, and Meghalaya, in which Christianity is the largest religion. In Brazil the variation in cross-gender friending broadly aligns with spatial patterns of socioeconomic status. The wealthier regions in the south exhibit much more gender segregation compared to the poorer regions in northern Brazil.

To explore whether the within-country variation in CGFR corresponds to meaningful variation in gender relations and attitudes, we study whether responses to several gender-related questions in the World Value Survey (WVS) vary systematically with the CGFR in the region where the respondent is located. The WVS asks respondents whether they agree with various statements related to gender (see Haerpfer et al., 2022). We focus on respondents' attitudes towards women's rights for post-secondary education, women's right to work, gender equality in a democracy, and whether women make good political leaders.

Table 1 shows estimates from regressions of WVS respondents' attitudes towards gender issues against the CGFR. Each column regresses within-country normalized responses to a given statement against the within-country normalized value of the $CGFR_c^{200}$. We control for a rich set of demographic characteristics which typically correlate with attitudes towards women, including religion, education, and gender, all interacted with country fixed effects.

Respondents from areas with less social gender segregation are typically more sup-

Table 2—Cross-Gender Friending Ratio and U.S. County-Level Characteristics

	CGFR (Top-200 Friendships)			
Share White	-0.153***	-0.116***	-0.135***	-0.150***
	(0.015)	(0.016)	(0.016)	(0.015)
Share Black	0.079***	0.116***	0.064***	0.060**
	(0.021)	(0.020)	(0.023)	(0.025)
% Adhering to any Religion		-0.004	-0.003	0.013
		(0.014)	(0.014)	(0.008)
Religious Congregations (Per 1000 Residents, Log)		-0.018***	-0.024***	-0.010***
		(0.003)	(0.006)	(0.003)
Median Household Income (Log)			-0.061***	-0.066***
			(0.011)	(0.006)
Female LFPR (Ages 25 to 64)			0.167***	0.167***
			(0.043)	(0.022)
% Female with College Degree			0.041	0.089***
			(0.037)	(0.022)
State FE				Yes
Observations	3,134	3,132	3,132	3,132
R^2	0.63	0.67	0.72	0.82

Note: Each column shows a regression of county CGFR against county-level covariates. Data on demographic composition, median household income, labor force participation, and educational attainment comes from the 2022 5-Year American Community Survey. Estimates of religious congregation density and religious adherents by county are provided by the US Religious Census. A religious adherent is defined as any person who claims to practice a religion. Regressions are weighted by county population in the 2020 Census. Standard errors clustered by state appear in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1.

portive of equal treatment of men and women across a range of domains than respondents from areas of the same country with more social gender segregation. For example, column 1 of Table 1 shows that living in a region with a one within-country standard deviation higher CGFR corresponds to a 0.133 within-country standard deviation-lower agreement with the belief that university is more important for boys than for girls.

Columns 2 and 4 of Table 1 also show that respondents from high-CGFR locations are more likely to disagree with the viewpoint that men have a greater right to work and that they are better political leaders. Column 3 highlights that respondents from regions within a country with less gender segregation in friendships are more likely to agree that it is essential that women and men have the same rights in a democracy.

These results suggest that our subnational measures of the CFGR pick up meaningful regional variation in attitudes regarding women's rights and access to opportunity.

III. Cross-Gender Ties in the U.S.

In this final section, we explore how various county-level characteristics are correlated with gender segregation in friendship networks within the United States. We test three categories of correlates: racial composition, religious adherence, and socioeconomic characteristics using data from the ACS (US Census Bureau, 2022).

Table 2 presents regressions of county-level observables against the CGFR. About 63% of the cross-county variation in the CGFR can be explained by county-level racial composition, and the effect size remains robust to controlling for other demographic characteristics, religious adherence, and state fixed effects. A increase in the share of White people of 10% corresponds to an decrease in the

CGFR of between 0.012 and 0.015. Conditional on the share of the population that is White, an increase in the share of the population that is Black is associated with lower degrees of gender segregation in networks. Given that the middle 95% of counties have a CGFR between 0.55 and 0.67, these magnitudes represent large within-country shifts in social gender segregation. The additional covariates included in column 3 of Table 2 only explain an additional 9% of the variation in the CGFR, with state fixed effects explaining an additional 10% (column 4).

IV. Conclusion

In this paper, we highlight how social media data can be used to measure gender segregation in social networks at scale (see Kuchler and Stroebel, 2023). The metrics that we produce allow for the comparison of such differences both across countries and at a fine level within countries. We hope that this data will open new possibilities for research on both the causes and effects of gender segregation in social networks.

REFERENCES

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

Bailey, Michael, Drew Johnston, Theresa Kuchler, Dominic Russel, Bogdan State, and Johannes Stroebel. 2020. "The determinants of social connectedness in Europe." 1–14, Springer.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018. "Social connectedness: Measurement, determinants, and effects." *Journal of Economic Perspectives*, 32(3): 259–280.

Bailey, Mike, Drew Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel. 2024. "The Social Integration of International Migrants: Evidence from the Networks of Syrians in Germany." *Journal of Political Economy*.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, et al. 2022*a*. "Social capital I: measurement and associations with economic mobility." *Nature*, 608(7921): 108–121.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, et al. 2022*b*. "Social capital II: determinants of economic connectedness." *Nature*, 608(7921): 122–134.

Currarini, Sergio, Matthew O Jackson, and Paolo Pin. 2009. "An economic model of friendship: Homophily, minorities, and segregation." *Econometrica*, 77(4): 1003–1045.

Haerpfer, Christian, Ronald Inglehart, Alejandro Moreno, Christian Welzel, et al. 2022. "World values survey: Round seven-country-pooled datafile version 5.0." *JD Systems Institute & WVSA Secretariat*, 12(10): 8.

Kuchler, Theresa, and Johannes Stroebel. 2023. "Social Interactions, Resilience, and Access to Economic Opportunity: A Research Agenda for the Field of Computational Social Science." *Handbook of Computational Social Science for Policy*, 405–419.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual review of sociology*, 27(1): 415–444.

UNDP. 2024. *Human Development Report 2023-24: Breaking the Gridlock: Reimagining Cooperation in a Polarized World.*

US Census Bureau. 2022. "American Community Survey 5-Year Estimates."