

ANALYZING FACEBOOK DATA TO UNDERSTAND REGIONAL CONNECTIVITY

Wijeratne, Yudhanjaya ;Samarajiva, Rohan;

Lokanathan, Sriganesh;Surendra, Aparna;Fernando, Lasantha;

© 2018, LIRNEASIA



This work is licensed under the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted use, distribution, and reproduction, provided the original work is properly credited.

Cette œuvre est mise à disposition selon les termes de la licence Creative Commons Attribution (<https://creativecommons.org/licenses/by/4.0/legalcode>), qui permet l'utilisation, la distribution et la reproduction sans restriction, pourvu que le mérite de la création originale soit adéquatement reconnu.

IDRC Grant/ Subvention du CRDI: 108008-001-Leveraging Mobile Network Big Data for Developmental Policy

1.1. Annex 15: Analyzing Facebook Data to understand Regional Connectivity

Yudhanjaya Wijeratne, Rohan Samarajiva, with contributions from Sriganesh Lokanathan and research support from Aparna Surendra and Lasantha Fernando

An ongoing study using social network data aims to understand social linkages between countries, and the corresponding links in economic activity and human movement.

In July 2017, Bailey et al published a paper titled Measuring Social Connectedness. In it, they introduced a new measure of social connectedness between U.S. county-pairs, as well as between U.S. counties and foreign countries. The measure (called the "Social Connectedness Index", or SCI), was based on the number of active Facebook friendship links between geographies as of April 2016, using IP to allocate people to countries and tabulating the number of connections between them.

Their study found that the Facebook data, for US counties, correlated strongly with trade, migration and facets of geography. The Facebook study also suggested that the SCI was detecting facets of homophily, showing correlations between friendships links and distance, and shared histories of migration between US counties.

Potent arguments for the existence of such networks in trade and migration are found elsewhere as well: Filip Gariz's research shows the links between homophily and migration, and Başak Bilecena, Markus Gamperb, Miranda J.Lubbers in "*The missing link: Social network analysis in migration and transnationalism*" point out the role that social networks in analyzing migration.

Clusters or networks of economies? A macroeconomy study through GDP fluctuation correlations by M. Ausloos and R. Lambiotte also introduced methods of correlating GDP fluctuations and proved that a economies can be expressed in terms of interlinked networks and societies.

Others, such as Trevor Martin, analyzing such networks in <http://www.shorttails.io/are-worldwide-economies-correlated/>, found strong correlations between GDP growth over time between certain nations and blocs, indicating network effects that may not be readily visible without specialized political knowledge.

Thus goal of our ongoing study is to examine existing coalitions of countries in the light of this social connectivity, using it to determine qualities such as the interconnectedness of a region and to analyze the social organization of countries within it, providing a replicable metric and methodology for doing so that relies on quantifiable data as opposed to varying social and political insight. A final goal is to attempt to use this data to identify communities of nations based on their social and economic connectivity.

FINDINGS

Our exploratory analysis had revealed that this social network data does strongly correlate to trade and migration between countries, but not to the distance between countries.

Figure 01: gross bilateral migration from 2013 versus friend links between country pairs.

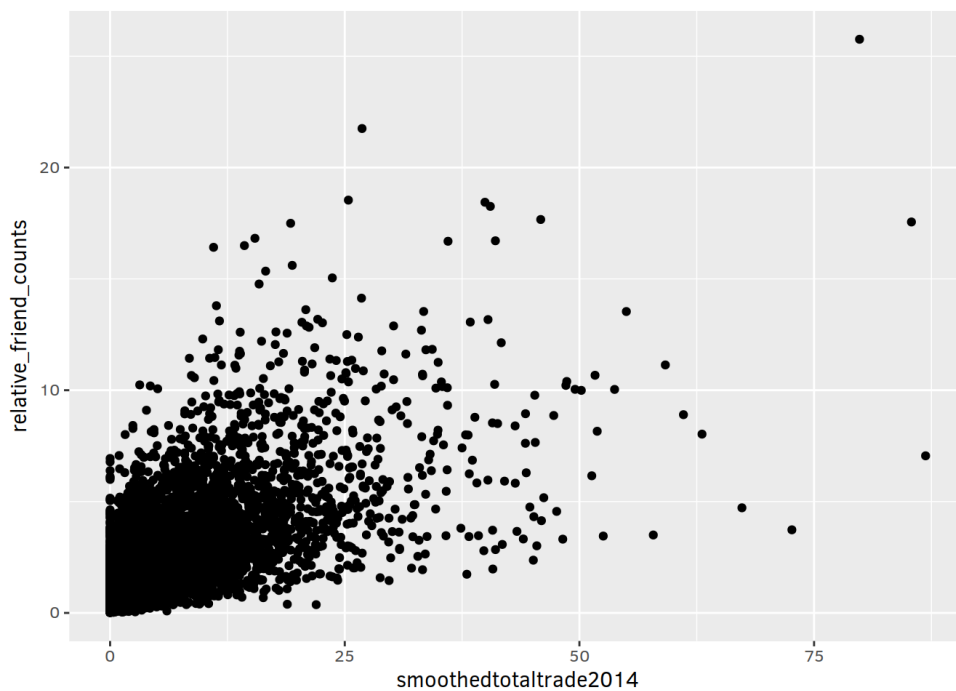
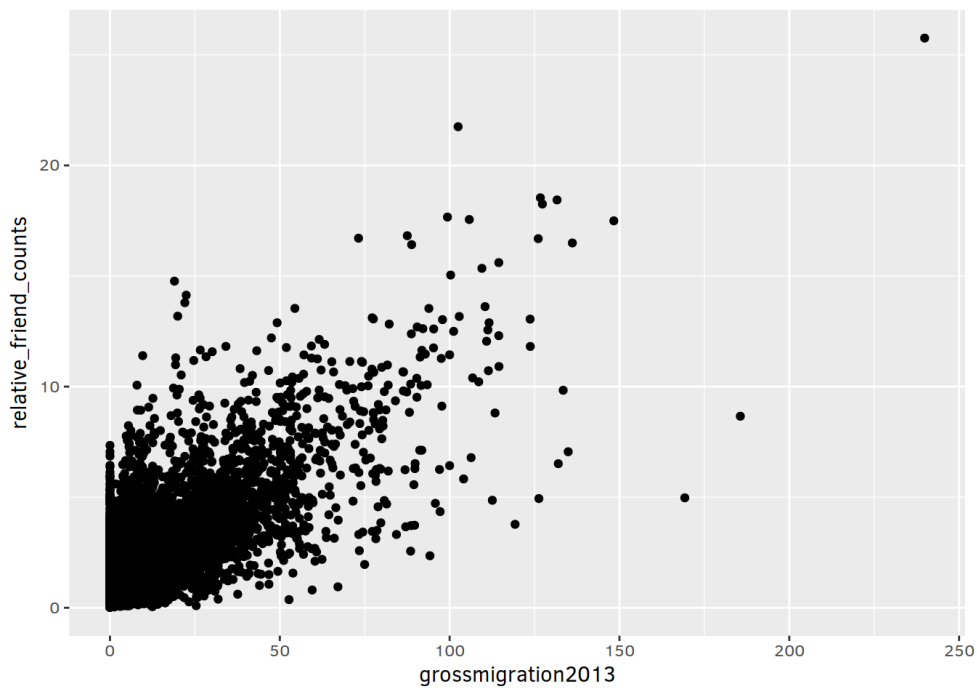
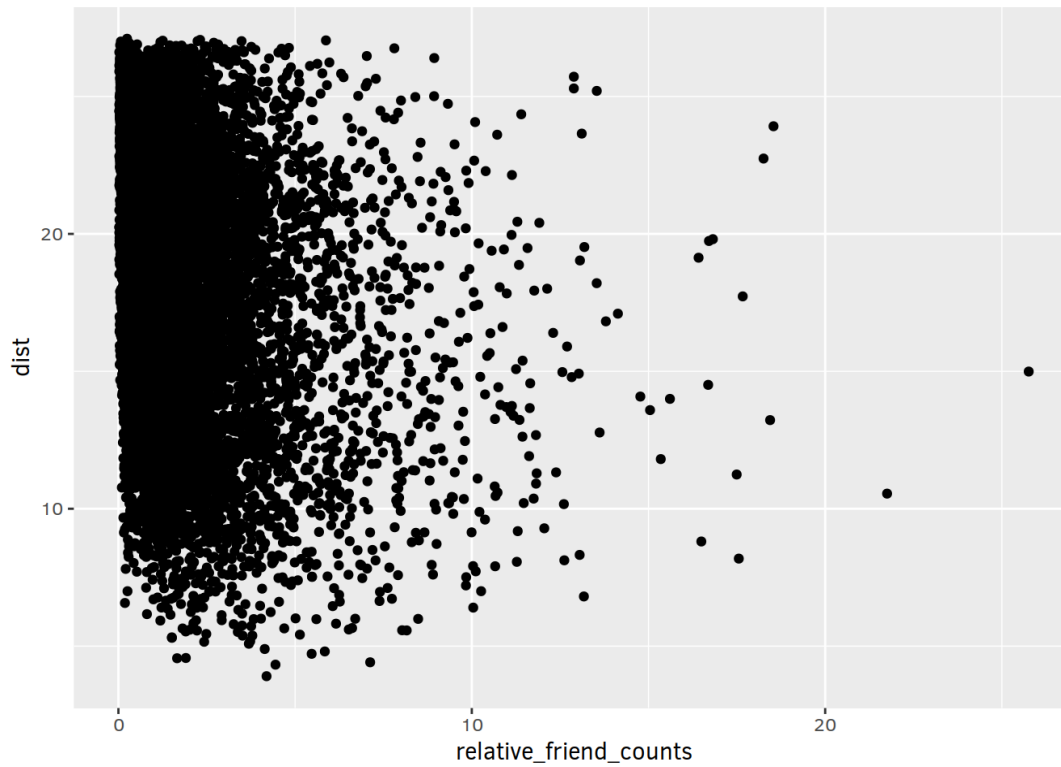


Figure 02: gross bilateral trade from 2014 versus friend links between country pairs.

Linear regression models built for these three factors - using connectivity versus trade and migration together - show an R-squared value of 0.6443, indicating that Facebook connectivity can be used as an able proxy for a combination of these two factors. No coherent and consistent relationship has been found between distances between nations

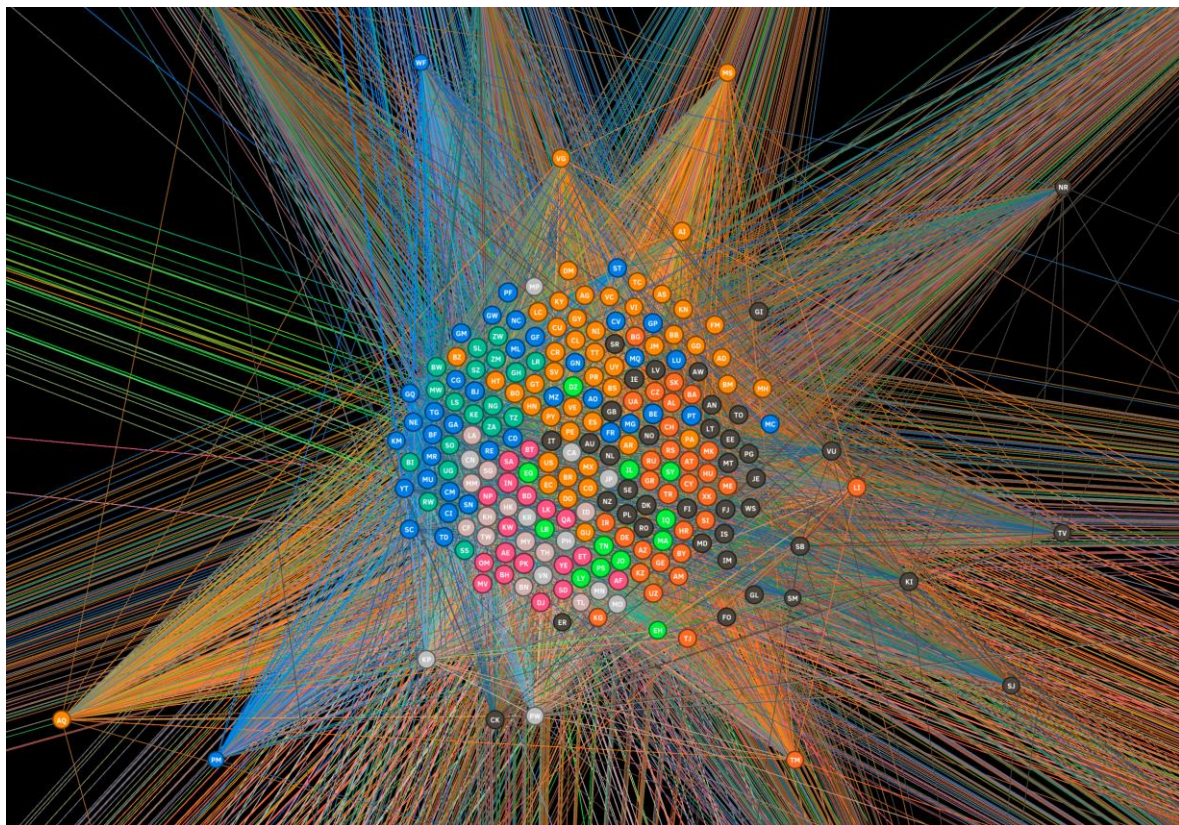
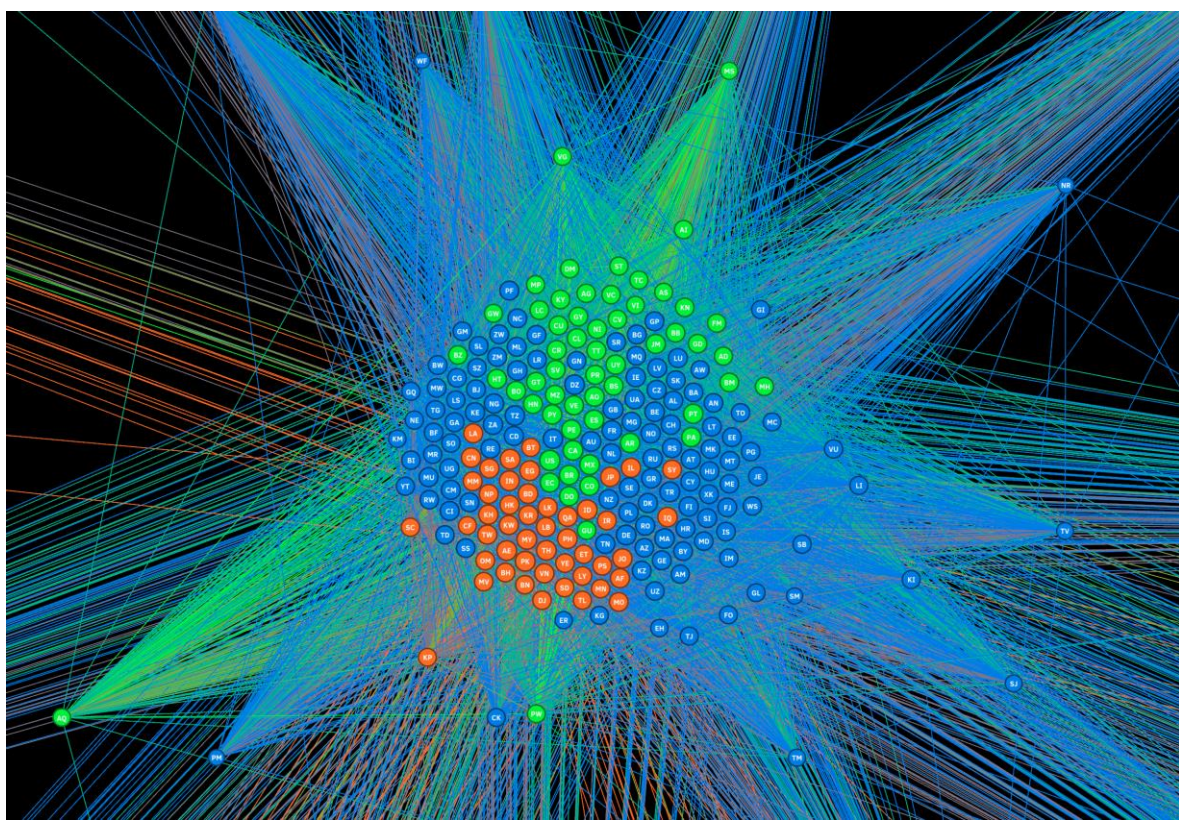
and their social connectivity.

Figure 03: distances between countries and friend links between them



Thus, social network connectivity between nations potentially serving as a possible way of exploring networks of international communities, human movement and economic activity. This often appears to transcend governmental and political barriers (such as between India and Pakistan, for example).

Thus this data, when analyzed worldwide with known community detection methods, reveals clusters of countries that are socially and economically integrated. The following maps have been constructed with the Louvain method at a range of resolutions: at higher resolutions the world is seen to split into three communities, and at lower, as many as nine, or seventeen.



We wish to explore these relationships further, both by examining homophily (as expressed in the seminal Birds of a Feather paper) at large with this data, and by also analyzing various existing political blocs such as the SAARC, ASEAN, BIMSTEC and others to understand how best this new metric can be put to use, and also by using different algorithms for community detection to yield better maps of the world.

The lack of correlation with geography indicates that this network may not follow the geography of the world, but rather may yield new insight into how countries form socioeconomic communities.

CHALLENGES

- **Facebook data is not accurately representative of regions such as Russia, Japan and China, where alternate networks are popular**
We will need to come up with a weighting variable based on Facebook penetration statistics, or factor in data from other networks
- **Data pre-processing done by Facebook**
What we have is not the raw data for the counts of users, but an aggregated matrix of normalized connectivity on a country level. It is difficult to reverse-engineer back to the original population stats. The network could be understood much better with the raw data, but it is unlikely that this can be obtained
- **Alternate patterns of Facebook usage from certain parts of the world**
One of the underlying assumptions made here is that the Facebook data is representative of Facebook users. This is so in the US, as proven in the original paper, but new research from Lirneasia in Myanmar using a non-representative sample showed users with up to twelve profiles interacting with thousands of others with all such profiles. This may be just an isolated case in Myanmar, with its unique political and cultural situation, or may be a pattern at large across Asia: more representative research will be needed here.
- **Missing data and lack of precise trade and migration data**
Trade and migration datasets miss out on illegal economic activity and human movement. Our data is from the World Bank and the Correlates of War project; it would be ideal to have some imputation of expected trade and migration, but this data needs to be found or build manually after studying historical migration and trade worldwide.