

Countries of a Feather

Analyzing Homophily and
Connectivity Between Nations Through Facebook Data

Yudhanjaya Wijeratne, Sriganesh Lokanathan, Rohan Samarajiva
TPRC 46
Washington DC, October, 2018



This work was carried out with the aid of a grant from the International Development Research Centre, Canada and the Department for International Development UK..



Introduction

In 2017, two studies put forth some interesting facts about Facebook:

Barnett and Benfield (2017):

- Facebook friend links are highly correlated with bilateral migration between nations (2013)
- Facebook friend links also exhibit some facets of homophily, as defined by McPherson et al (2001).
- Using a highly aggregated snapshot of the Facebook network from 2012.

Bailey et al. (2017):

- Facebook friend links correspond with historical migration and trade patterns between states in the US
- Facebook friend links also exhibit some facets of homophily, as defined by McPherson et al (2001).
- Using a snapshot of the Facebook network from 2016 (for the United States).

Why is this important?

Both studies seem to indicate that Facebook friend links can be of value to the study of two seemingly related concepts:

International connectivity

- Poorly defined, but of great importance to policy
(Oxera Consulting, 2010; Oxford Economics, 2013; Arvis and Shepherd, 2013)
- Traditionally studied through trade and migration
(McNeill and Adams, 1978; Hont, 2005; Egger, von Ehrlich and Nelson, 2012; Castles, Miller and Haas, 2014; Leonard, 2016)
- Recent literature acknowledges the need to study ICT-based communication networks
(Gould and Panterov, 2017; Gould, Kennet and Panterov, 2018)

Homophily

- Better defined - a universal principle in how humans form ties
McPherson et al. (2001)
- Traditionally studied via facets like shared languages, geographical proximity measures, shared value systems etc.
- Rich recent history of literature exploring the existence of various facets in social media, trade and international politics
(Haythornthwaite, 2005; boyd and Ellison, 2007; Pempek et al, 2009; Mislove et al., 2010; Zhou, 2011, Maoz, 2012)

Research questions

We received data from Facebook – friend links between nations, for 2016, in a schema more granular than that available to Barnett and Benfield. We were able to conduct a simple exploratory analysis.

1. How does this more recent data correlate with international migration and trade?
2. Can we detect factors of homophily in the Facebook data?
3. What does this tell us about the value of the Facebook friend network and/or homophily to analysis of international connectivity?

Data

Facebook friend links

Undirected network listing 259 countries along with a value for the number of friend links between each pair of countries, as per the schema devised by Bailey et al. (2017).

Facebook friend links between users over 18 years of age are counted and all users and links aggregated to country levels. The counts of friend links are then divided by that of the highest country pair. Some linear transformation has been performed, either for obfuscation or ease of interpretation.

Gross migrant stock

Directed network. Migrant stock data for 2013, 2015 and 2017 from the United Nations database, POP/DB/MIG/Stock/R ev.2017. Processed into gross migrant stock as the sum of migrant stock from two countries in each other.

2013 data covers 10041 country pairs - 230 unique reporters and 227 partners. 2015 data is between 231 reporters and 228 partners, for a total of 8195 country pairs. 2017 data is between 230 reporters and 228 partners for 8241 country pairs.

Gross goods trade

Directed network. 2015, 2016 and 2017 Bilateral goods trade flow from the DESA/UNSD, United Nations Comtrade database (2018). Processed similar to migrant stock.

2015 goods trade data covers 17014 country pairs - 155 unique reporters and 235 partners. 2016 data is between 147 reporters, 225 partners, for a total of 16735 country pairs. 2017 data is between 89 reporters and 225 partners for 12581 country pairs.

Gross services trade

Directed network. Bilateral services trade data for 2013, 2015 and 2016 from the World Bank Trade in Services Database (2018).

2013 services trade data covers 4281 country pairs - 41 unique reporters and 234 partners. 2015 data is between 39 reporters and 234 partners, for a total of 4259 country pairs. 2016 data is between 26 reporters and 37 partners for 561 country pairs.

Data – II

For homophily, we used data from CEPII'S Gravity dataset (Head, K., & Mayer, T.,2014)

1. Common borders (expressed as geographical contiguity)
2. Common official language
3. Common language spoken by over 9% of the population
4. The current or prior existence of a colonial relationship
5. The existence of a current colonial relationship
6. The existence of a common currency
7. The existence of a Free Trade Agreement (as defined by the World Trade Organization).

The selection of these values is based loosely on their value in representing three facets of homophily posited by McPherson et al.(2001): geographical closeness, shared language, and shared values

Data – III

| reporter | partner | Facebook_friend_counts | gross_migrant_stock_2013 | gross_migrant_stock_2015 | gross_migrant_stock_2017 | gross_goods_trade_2015 | gross_goods_trade_2016 | gross_goods_trade_2017 | gross_services_2013 | gross_services_2014 | gross_services_2015 | gross_services_2016 |
|----------|---------|------------------------|--------------------------|--------------------------|--------------------------|------------------------|------------------------|------------------------|---------------------|---------------------|---------------------|---------------------|
| USA | MEX | 17092.95 | 13799404 | 12926559 | 13582377 | 1.03E+12 | 1.01E+12 | 1.08E+12 | 204498.1 | 216835.3 | 230596.9 | 0 |
| MEX | USA | 17092.91 | 13799404 | 12926559 | 13582377 | 1.03E+12 | 1.01E+12 | 1.08E+12 | 204498.1 | 216835.3 | 230596.9 | 0 |
| MYS | IDN | 10291.79 | 1053206 | 1072634 | 1094156 | 3.16E+10 | 2.81E+10 | 3.43E+10 | 0 | 0 | 0 | 0 |
| IDN | MYS | 10291.79 | 1053206 | 1072634 | 1094156 | 3.16E+10 | 2.81E+10 | 3.43E+10 | 0 | 0 | 0 | 0 |
| USA | PHL | 6370.36 | 2034711 | 1931595 | 2112936 | 3.5E+10 | 3.49E+10 | 3.75E+10 | 31108 | 33534 | 37236 | 0 |
| PHL | USA | 6370.307 | 2034711 | 1931595 | 2112936 | 3.5E+10 | 3.49E+10 | 3.75E+10 | 0 | 0 | 0 | 0 |
| ARE | IND | 6271.529 | 2864416 | 3511323 | 3322285 | 7.87E+10 | 7.92E+10 | 0 | 0 | 0 | 0 | 0 |
| IND | ARE | 6271.50248 | 2864416 | 3511323 | 3322285 | 7.87E+10 | 7.92E+10 | 3.75E+10 | 0 | 0 | 0 | 0 |
| USA | IND | 6084.176 | 2063543 | 1972008 | 2310603 | 1.29E+11 | 1.32E+11 | 1.30E+11 | 166955 | 184753 | 211033 | 0 |
| IND | USA | 6084.176 | 2063543 | 1972008 | 2310603 | 1.29E+11 | 1.32E+11 | 1.30E+11 | 0 | 0 | 0 | 0 |
| GBR | USA | 5514.072 | 981120 | 927149 | 938697 | 2.42E+11 | 2.29E+11 | 2.28E+11 | 797760.9 | 848839.2 | 854413.7 | 455364.3 |
| USA | GBR | 5514.05171 | 981120 | 927149 | 938697 | 2.42E+11 | 2.29E+11 | 2.28E+11 | 797760.9 | 848839.2 | 854413.7 | 0 |
| CAN | USA | 5410.767 | 1184060 | 1185427 | 1199825 | 1.12E+12 | 1.06E+12 | 1.13E+12 | 1122732 | 1098005 | 826768.3 | 0 |
| USA | CAN | 5410.76528 | 1184060 | 1185427 | 1199825 | 1.12E+12 | 1.06E+12 | 1.13E+12 | 1122732 | 1098005 | 826768.3 | 0 |
| IND | BGD | 5355.72669 | 3263820 | 3205453 | 3174561 | 1.26E+10 | 6.35E+09 | 6.27E+09 | 0 | 0 | 0 | 0 |
| BGD | IND | 5355.698 | 3263820 | 3205453 | 3174561 | 1.26E+10 | 0 | 0 | 0 | 0 | 0 | 0 |
| SAU | PHL | 4760.466 | 1029540 | 488900 | 584741 | 2.24E+09 | 1.48E+09 | 0 | 0 | 0 | 0 | 0 |
| PHL | SAU | 4760.451 | 1029540 | 488900 | 584741 | 2.24E+09 | 1.48E+09 | 1.35E+09 | 0 | 0 | 0 | 0 |

All distributions herein are nonparametric.

Findings - I

Facebook friend links between pairs of countries show strong positive correlations with the gross migrant stock, gross services trade and gross goods trade.

| DATA | | | | | ANALYSIS | | |
|----------------------|------|-----------|----------|---------------|----------------|---------------|-----------|
| | Year | Reporters | Partners | Country pairs | Spearman's rho | Kendall's tau | p-value |
| Gross migrant stock | 2013 | 230 | 227 | 10041 | 0.729 | 0.570 | < 2.2e-16 |
| | 2015 | 231 | 228 | 8195 | 0.741 | 0.576 | < 2.2e-16 |
| | 2017 | 230 | 228 | 8241 | 0.736 | 0.571 | < 2.2e-16 |
| Gross services trade | 2013 | 41 | 234 | 4281 | 0.756 | 0.579 | < 2.2e-16 |
| | 2015 | 39 | 234 | 4259 | 0.766 | 0.589 | < 2.2e-16 |
| | 2016 | 26 | 37 | 561 | 0.771 | 0.581 | < 2.2e-16 |
| Gross goods trade | 2015 | 155 | 225 | 17014 | 0.702 | 0.547 | < 2.2e-16 |
| | 2016 | 147 | 225 | 16735 | 0.705 | 0.549 | < 2.2e-16 |
| | 2017 | 89 | 225 | 12851 | 0.737 | 0.577 | < 2.2e-16 |

Findings – II

This similarity seems to be shared mostly with gross migrant stock.

- 1) Degree and betweenness centrality nodes in each network (a rudimentary way of examining network structure) show a tendency for Facebook and migration to share more countries in the top 20.
- 2) Linear regression models using Facebook friend links can predict 47% of the variation of migration data.

Overall, this adds more evidence to Barnett and Benfield (2017) and Bailey et al (2017)'s findings.

Findings – III

Despite us using data from a social network proven to exhibit homophily, we perceived only very weak positive correlations between Facebook friend links and homophilic factors.

1. Common borders (0.187)
2. The existence of a common language spoken by over 9% of the population (0.11)
3. The existence of a Free Trade Agreement (0.1)
4. The sharing of a colonial relationship (0.1).

While statistically significant, these relationships had very low predictive capability.

No combination of these achieved an R^2 of over 0.06 in a linear model.

Conclusions

This presents further evidence that Facebook friend links mirror the movements of humans and capital – traditional metrics of international connectivity.

1. These relationships exist not just for 2012 (Barnett and Benfield, 2017) or just for the United States (Bailey et al, 2017)
2. Given that these are be citizen-driven links, formed independently of government lead relations (Barnett and Benfield, 2017) from a network representing over 2 billion people worldwide, this may be of some use in the study of international connectivity – especially what in IR is termed “people-to-people relations”.
3. Given the similarities to migrant stock networks, Facebook data may, with further study, be used to cross-validate assumptions about international connectivity based on migration data known to be flawed.
4. The effects of homophily exist in social networks, but certain traditional facets (such as geography and spoken languages) may be weakening. The old metrics may not be useful to international connectivity. Or this could be an artifact of the Facebook network and not a useful finding.

However

This study is only an exploratory analysis. Further research / data manipulation to be done to make this useful:

a) Facebook is not uniformly representative. Under-represented relationships (Russia, China etc) will either need to be scaled or removed, as some of these nations represent significant migrant stock and trading power. It is possible to normalize these links by weighing them against the ratio of the Facebook users in each country to the population of said country.

b) A key assumption is that Facebook links mirror real-world ones (Bailey et al, 2017). While substantial literature says so, the findings of Hurulle (2018), which showed users in Myanmar often creating multiple fake accounts to participate in sociopolitical discussions, show that these patterns may not hold true for some countries.

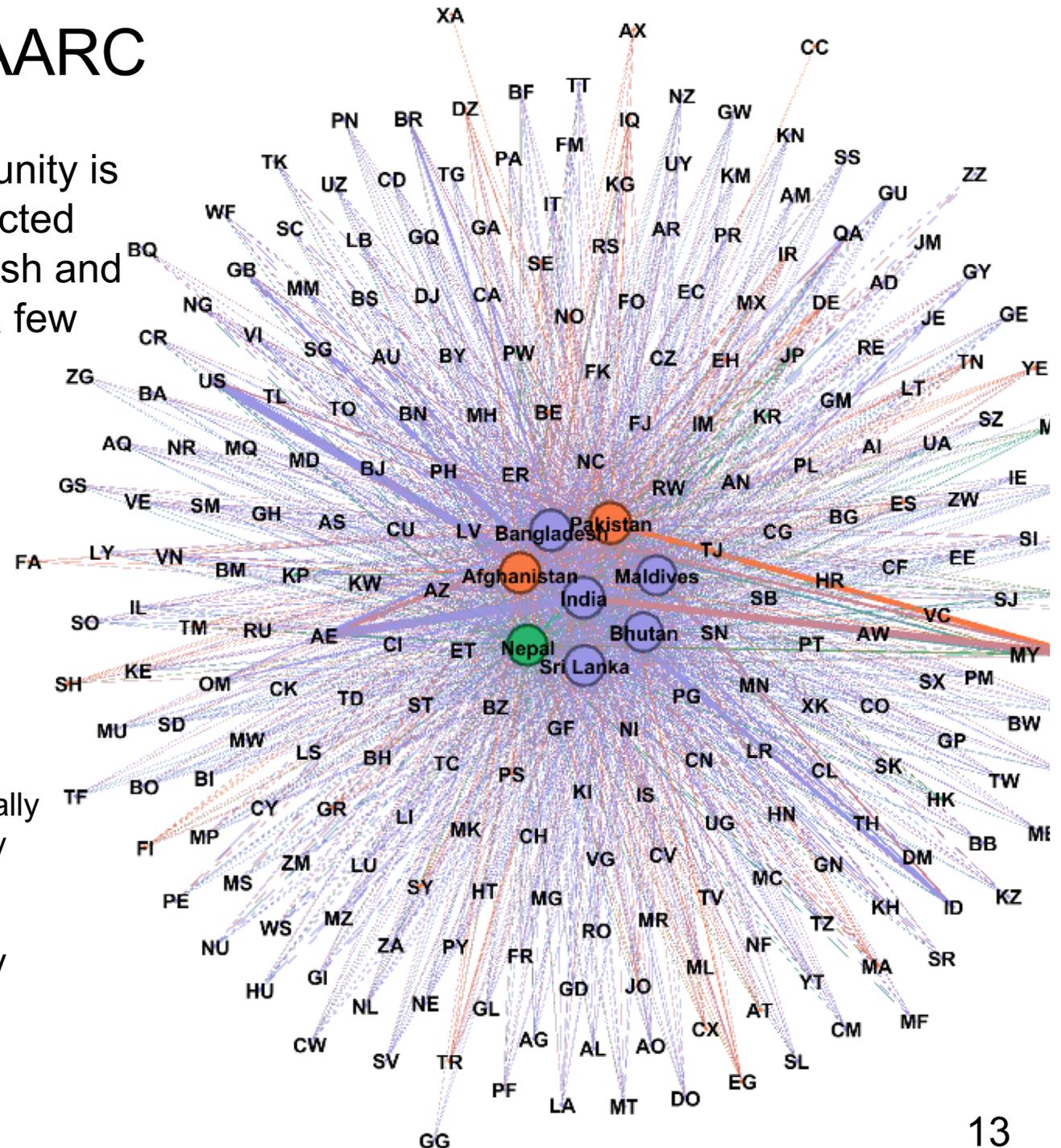
Applications I: SAARC

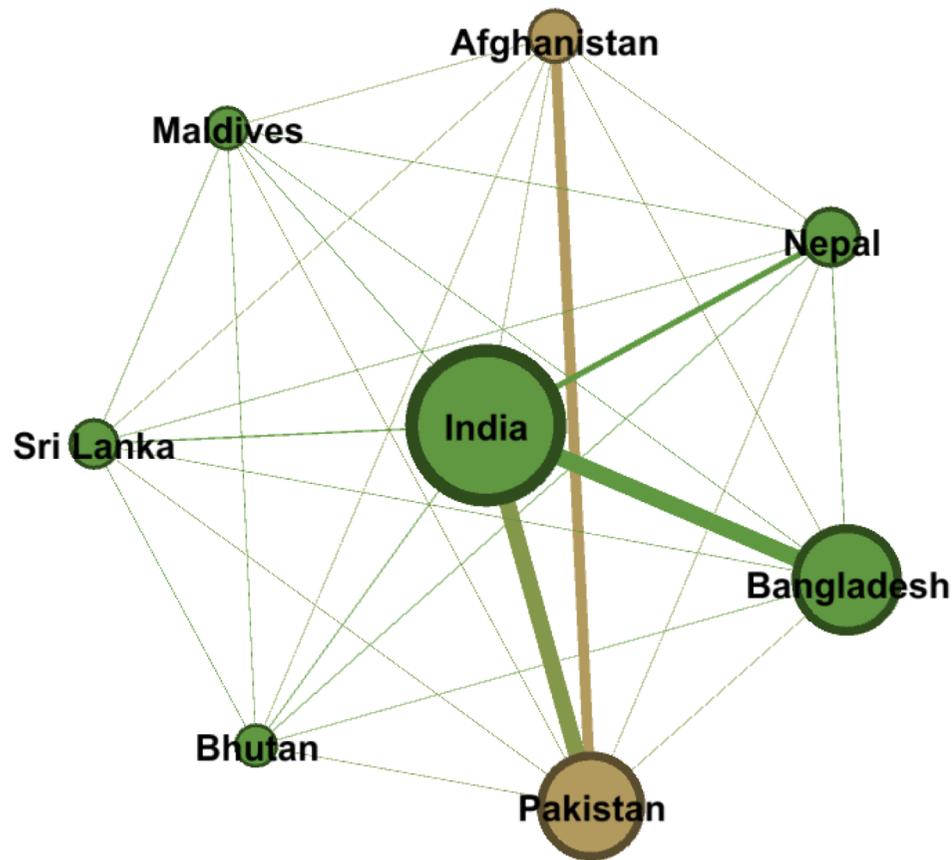
The SAARC 'social' community is actually three highly connected countries - India, Bangladesh and Pakistan - surrounded by a few other hangers-on.

| | |
|-------------|---------|
| India | 1822.28 |
| Bangladesh | 1047.43 |
| Pakistan | 1013.42 |
| Nepal | 282.52 |
| Afghanistan | 196.44 |
| Sri Lanka | 135.69 |
| Maldives | 35.40 |
| Bhutan | 24.05 |

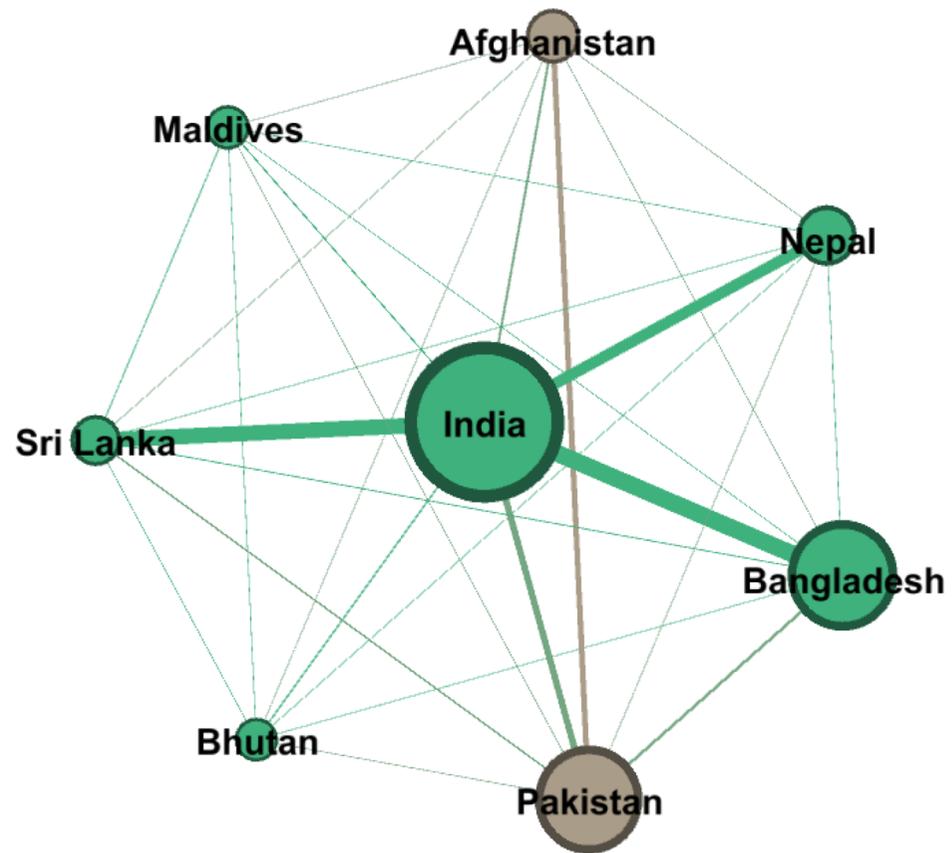
Pakistan and Afghanistan are socially connected to a different community altogether, as is Nepal.

(Algorithm: Louvain Method by Blondel et al, 2008.)





Gross migrant stock, 2015



Gross trade (goods), 2015

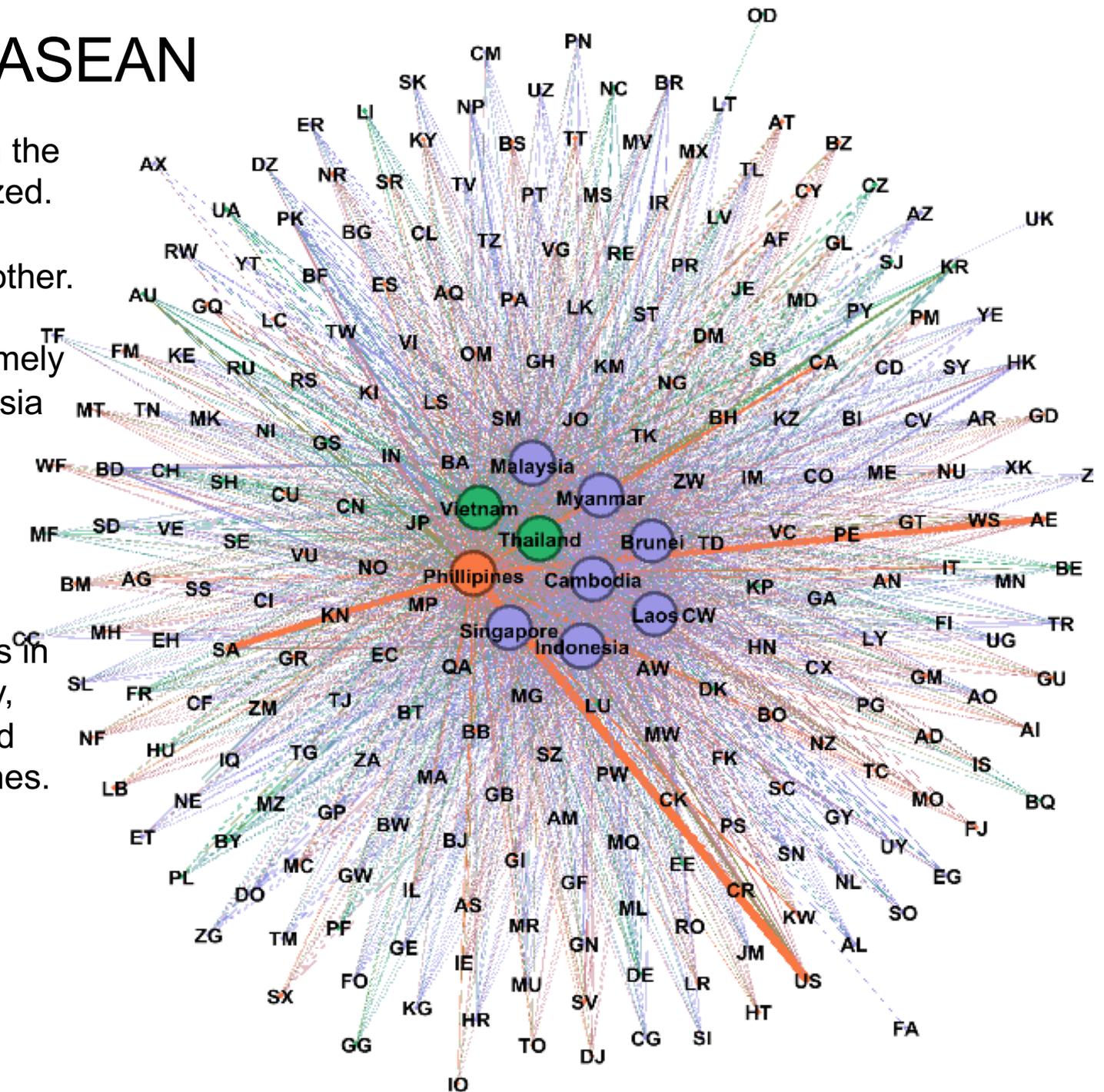
This same split appears in migration and trade. As in the social network, India remains the biggest influencer, and India, Bangladesh, Nepal, Sri Lanka, Bhutan and Maldives operate as one community, but Pakistan and Afghanistan have their own little sub-community with migrant stock and trade.

Applications II: ASEAN

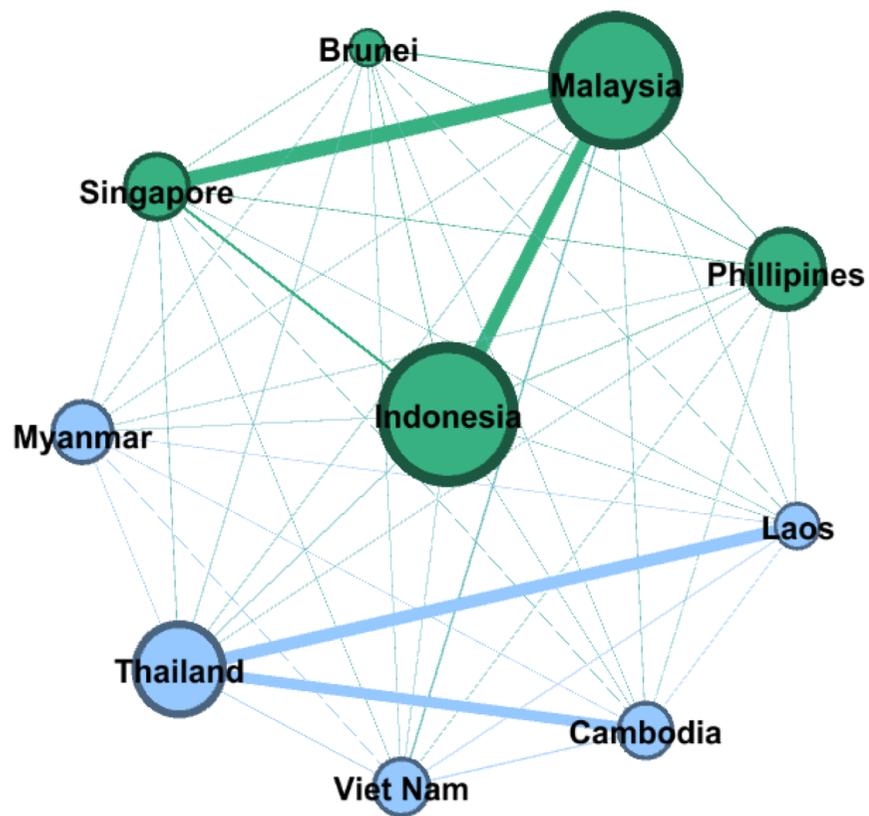
The ASEAN community, on the other hand, is less centralized. ASEAN countries are more evenly connected to each other.

The exception is one extremely strong link between Indonesia and Malaysia.

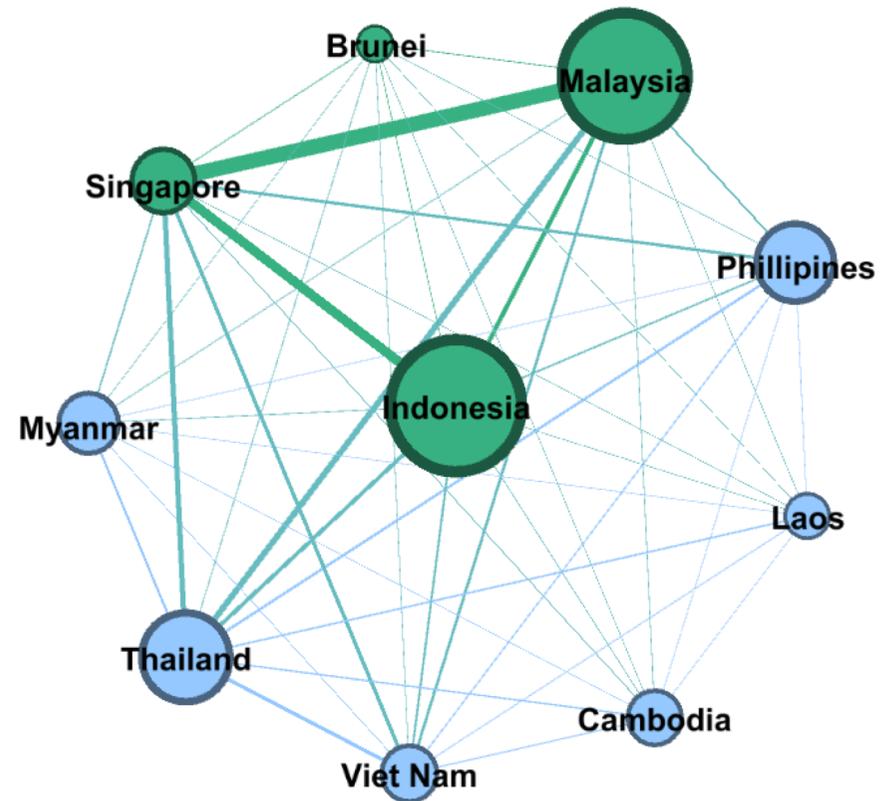
Most of the ASEAN belongs in the same global community, Vietnam and Thailand stand apart. As does the Philippines.



Data: FACEBOOK



Gross migrant stock, 2015

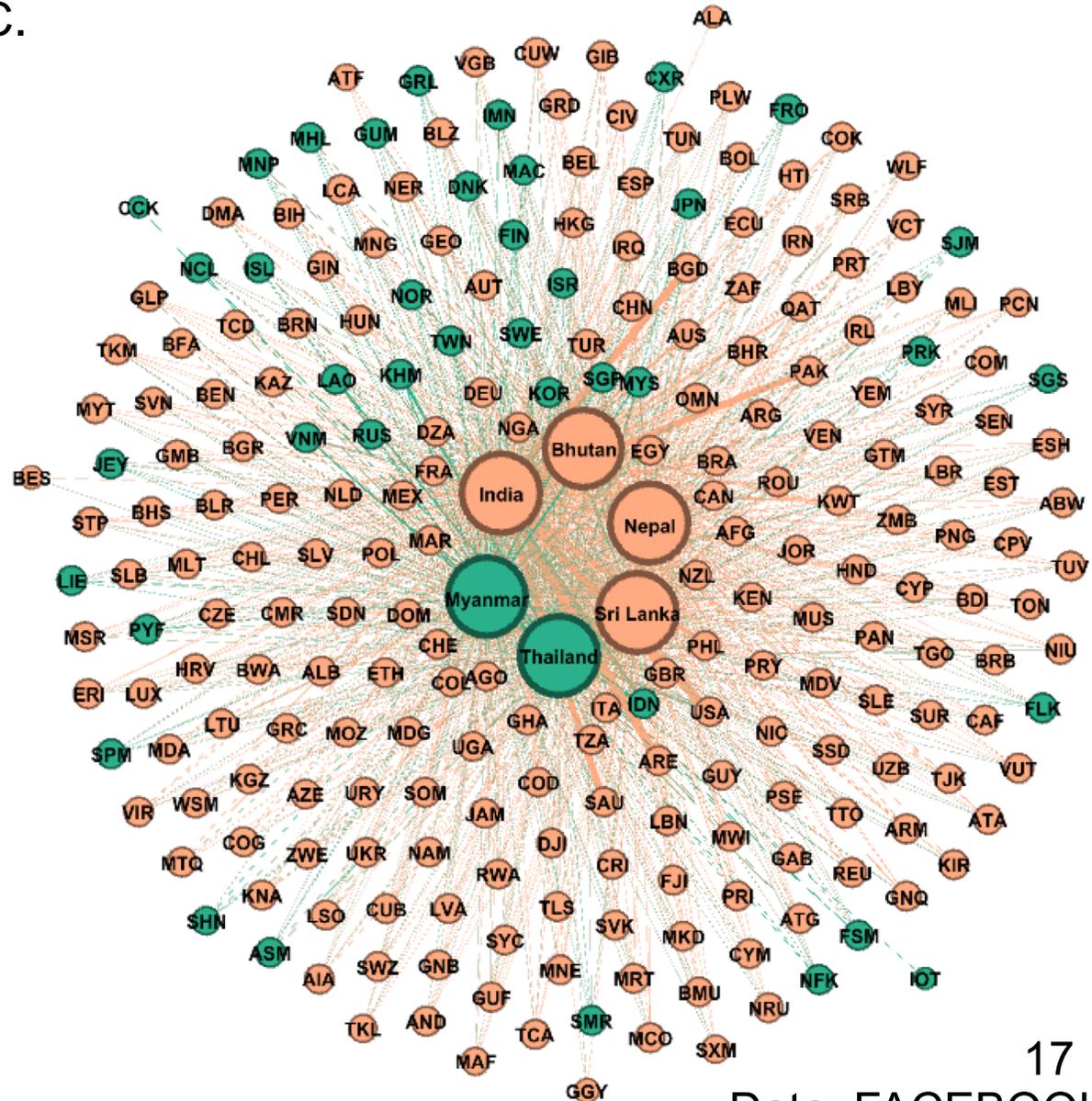
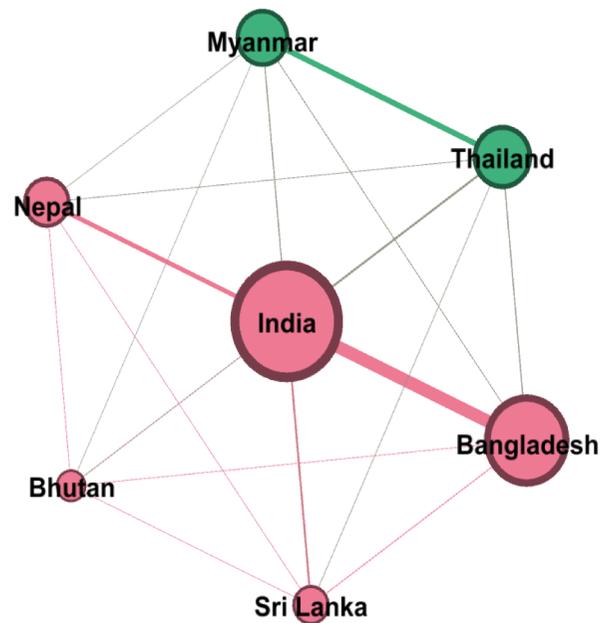


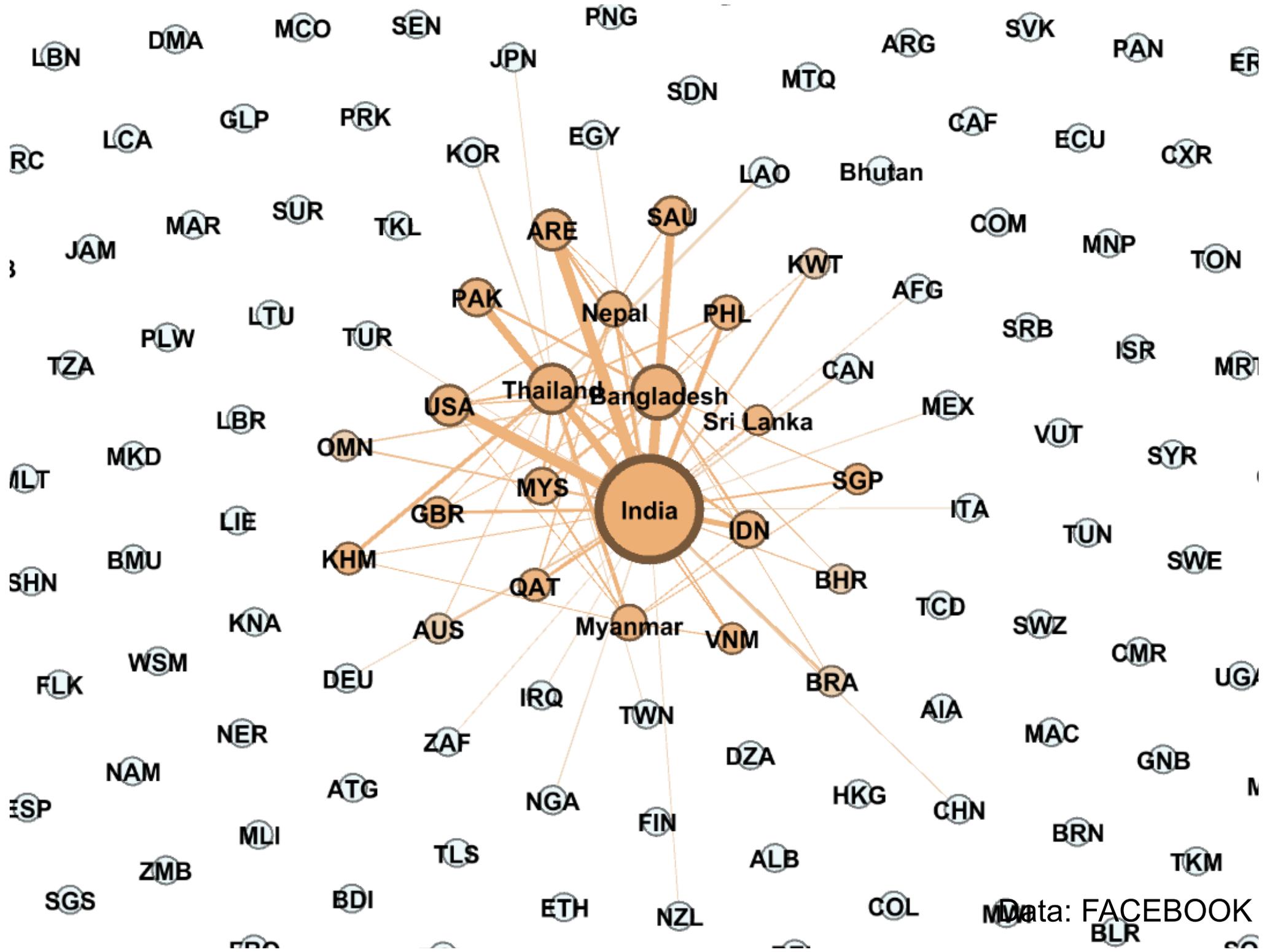
Gross trade (goods), 2015

To some extent this community split manifests itself in both migration and trade. Myanmar, Thailand, Viet Nam, Cambodia and Laos consistently sort themselves into the same community.

Applications III: BIMSTEC

Two parts of a tight sub-community from ASEAN mixed with five parts of a tight sub-community from SAARC.





Data: FACEBOOK

Caveats

- 1) Community detection this way is not an exact approach – different parameters of the same algorithm will yield different results
- 2) Results like these may correspond to “gut feel” and political/cultural context but are hard to rigorously quantify.
- 3) As stated before, representivity needs to be considered before using such analysis, especially since communities can be skewed by powerful nodes being misrepresented.