The connectivity of South Asian cities in infrastructure networks

Key words: betweenness centrality, communities, rail, road, information technology, air transport

Abstract: This map summarizes information on the connectivity of 67 important South Asian cities in infrastructure networks. The map combines four information layers to reveal a city's overall stature in the region's infrastructure networks, i.e. rail, road, air, and information technology networks. Three dimensions of connectivity are shown: edge thickness reflecting tie strength between pairs of cities; node size reflecting a city's betweenness centrality; and node color reflecting the dominant geographical orientation of a city's connections. A threshold is used for the edges to ensure the map does not appear clogged. The map shows that major connections tend to be within-country linkages between large cities. There are five communities in South Asia's urban infrastructure networks, which largely follow national borders. Delhi, Mumbai, Lahore, Karachi, Chennai, Colombo and Dhaka are shown to be important nodes for the infrastructural integration of South Asia, as these cities mediate flows between relatively unconnected communities and cities.

Introduction

The development of transport infrastructures has been shown to increase productivity, competition, and business activity because of the enhanced market access that comes with lowered transport costs (e.g. Sahoo and Sexena, 1999; Calderon and Serven, 2004; Straub et al., 2008). In general, having access to broad labor, resource and customer markets makes a location more attractive by providing productivity or profitability benefits in addition to having appealing unit costs for workforce and facilities operations. The economic impact of the deployment of transport infrastructures strongly depends on the resulting infrastructural connectivity of an economic entity, which is much more than the mere stock or quality of available infrastructures. Connectivity refers to the directness, geographical diversity, and density of an economic entity's infrastructure linkages with other entities in the network. For instance, a well-connected node in a rail network does not simply denote the presence of a large railway station. Rather, it refers to a railway station that has a large number and a wide variety of direct links and/or short-path indirect connections across the network, or is used by many other nodes to interconnect.

Given the scientific recognition and the policy perception that connectivity in transportation networks affects an economic entity's productivity and economic growth, a thorough understanding of the concept and the empirics of 'connectivity' is of major importance. In this map, we present an overview of the key features of the infrastructural connectivity of 67 major cities in South Asia. We construct an undirected and valued composite infrastructure network consisting 67x66/2 = 2.211 edges by aggregating four different data layers, i.e. rail, road, air, and information technology networks. The map reveals three empirical dimensions of a city's infrastructural connectivity in the South Asian context (for recent advances in mapping information-dense networks, see Hennemann, 2013): (1) a city's major connections, (2) the dominant geographical orientation of these connections, and (3) a city's role as a gateway for city-pairs that are not directly interconnected.

Data and methods

All South Asian cities with a population of more than 750.000 inhabitants in 2011 are included in the analysis, in addition to the capital cities Colombo (Sri Lanka), Malé (Maledives) and Thimphu (Bhutan) to ensure that all South Asian countries are represented. As connectivity infrastructures such as airports are often shared between cities located in close proximity, we aggregated cities located within a 50km range. For instance, although the Rawalpindi/Islamabad and Mumbai/New Bombay city-pairs initially featured as separate nodes meeting the selection threshold, we combined these into single units of analysis: population numbers are thereby aggregated, and the dominant city label is subsequently used (e.g., in what follows Mumbai also includes New Bombay and Thane).

The airline data layer is constructed around the number of direct weekly flights offered during the last week of May 2013. Data were mainly obtained through Google's web crawling service, and crosschecked with the SkyScanner passenger flight search engine (www.skyscanner.net) and data from the Official Airline Guide (OAG, www.oag.com). The strongest connection in this layer is Mumbai-Delhi (351 weekly flights), followed by Mumbai-Bangalore (202 weekly flights).

The road data layer is based on a network efficiency measure, which is computed by dividing anticipated travel time between cities by the Euclidean distance separating them. The anticipated travel time is drawn from Google Maps, after which 30mins are deducted to control for slow-going traffic in and out of city centers. The strongest connection is Rawalpindi-Peshawar along Pakistan's M1 Motorway (consisting of 6 lanes for the entire stretch), so that this connection is anticipated to take 1h56min for 173km thus leading to a very high efficiency of (116min-30min)/173km = 0,497min/km. This is followed by the Jalandhar-Ludhiana connection along the Grand Trunk Road in India at 0,53min/km. As can be gleaned from tour approach, longer but on average high quality inter-city linkages are deemed more connected than shorter but on average low quality inter-city connections.

The train data layer is constructed around the number of direct weekly trains offered during the last week of May 2013. Data were gathered from a variety of sources: for India indiarailinfo.com, for Bangladesh railway.gov.bd, for Pakistan pakrail.com, and for the few transnational links such as Lahore-Amritsar-Delhi and Kolkata-Dhaka we used a range of secondary sources. The strongest connection is Vadodara-Surat in India's Gujarat province (366 weekly trains), a section of the Indian rail network that is used by many trains connecting India's major cities (e.g. Jaipur-Mumbai and Delhi-Mumbai). The second strongest connection is Vadodara-Ahmadabad (318 weekly trains), located along roughly the same set of crucial train axes. Kabul, Thimphu, Malé and Colombo have no connectivity in this layer.

The Internet data layer is based on data garnered in the context of the DIMES project. DIMES is a distributed scientific research project that aims to study the structure and topology of the Internet, and results in what is by far the best data

source around for mapping day-to-day Internet geographies (Tranos and Nijkamp, 2013). The data are based on 'traceroute' measurements for 2010, which were made daily by a global network of more than 10.000 agents, voluntarily participating in this research project (for a description of the DIMES project, see Shavitt and Shir, 2005). DIMES volunteers derive the raw connectivity data through geo-locating Internet Protocol (IP) links: although 'Internet flows' are often thought to be 'immaterial', this is an infrastructural measure because IP links represent physical data links between city-pairs (these can be e-mails, file downloads, etc.). In this layer, each IP link represents a connection, whereby vertices are geo-coded at the level of the cities as defined in our framework. The strongest connection is Delhi-Bangalore (10.000 IP links), followed by Mumbai-Delhi (8,951 IP links).

We first logged measures in each of the layers to alleviate the skewness in the distributions. Applying the following transformation thereupon normalized the logged figures in each of the four layers: (original-min)/(max-min). All four data layers thus have a distribution between 0 (lowest connectivity) and 1 (maximum connectivity)[1]. Edges in the composite network are thereupon computed by taking the average score of the logged and normalized values in each of the different layers. The major connections in each of the four network layers are shown in figure 1[2].

Figure 1 about here

A fully connected network is not very appealing from an analytical point of view (Hennemann and Derudder, 2013). For instance, a fully connected network assumes there are no gateways, which is not a very realistic proposition in infrastructure networks. To circumvent this problem, which largely emanates from the fully connected road network, we impose a connectivity threshold to remove small and therefore conceptually not very meaningful connections: we set all edges in the composite network with a value <0,40 to 0. The resulting network has a much lower density of 0.128, but nonetheless a sizable QAP correlation of 0.827 with the original network (figure 2): this implies that the structure of composite network very closely follows the structure of the original network, but with a density that allows for a more thorough analysis of its topology and structure as only the meaningful links are retained.

Figure 2 about here

---

[1] For the road network, the formula is actually 1-(original-min)/(max-min) as low values in the raw data represent strong connectivity. After this transformation, larger values also represent stronger connectivity (in line with the other three layers).

[2] Note that for reasons of clarity, only the strongest connections are shown. For each layer, this implies a visualization threshold (not to be confused with the analytical threshold elaborated below) aimed at including roughly >10% of all 2211 possible connections, i.e. 258 rail connections (all transformed values >0.5 + all international connections), 218 for dimes (all values), 244 for flight (all values), and 321 for road (all road efficiency values >0.8 + major international linkages).

Finally, some manual adjustments were made, which implies retaining some values below the 0,40 threshold. This was done if (a) a city did not have 3 significant relations (in which case its 3 largest connections were included, e.g. for Thimphu), or if (b) specific transnational inter-city linkages, despite their relatively moderate value, represented conceptually significant linkages for the region's integration (e.g. the significance of the Lahore-Amritsar train and the Mumbai-Karachi flight for the infrastructural interconnection of India and Pakistan). The resulting dataset is used to map the major infrastructural inter-city connections in South Asia.

Two other connectivity measures drawn from this dataset – community membership and betweenness centrality – are used for plotting node color and node size, respectively.

Node color reflects the dominant geographical orientation of a city's connections. To this end, we applied a community detection algorithm, which divides a network in subnetworks in such a way that each subnetwork is densely connected internally. The community detection algorithm used here is the fast greedy modularity optimization algorithm for finding community structures (Clauset et al., 2004). The algorithm calculates the membership corresponding to the maximum modularity score, considering all possible community structures along the merges.

Node size reflects a city's betweenness centrality, which measures a city's role as a switching point for the interconnection of cities that are not directly connected. The betweenness centrality BC_a of a city a is given by the expression:

$$BC_a = \sum_{x \neq a \neq y} \frac{SP_{xy,a}}{SP_{xy}}$$

whereby SP_{xy} is the number of shortest paths from node x to node y, and $SP_{xy,a}$ is the number of these shortest paths running through a. However, as we are dealing with a weighted network, in practice links are considered in proportion to their capacity, which adds an extra dimension beyond the topological effects.

Conclusions

The strongest infrastructure linkages are, unsurprisingly, between the largest metropolises in South Asia. However, it is remarkable that, in contrast to other parts of the world, connectivity is very strongly impacted by national borders. Although border effects are visible across many different so-called 'global' transaction patterns (e.g. Hennemann et al., 2012; Taylor et al., 2013), these effects are so strong in this region that they define the basic geography. This pattern also surfaces in the other two indicators: communities closely follow national borders, and the few cities that act as transnational gateways have the strongest betweenness centrality.

There are 5 communities in South Asia's infrastructure networks: there is a community bringing together Pakistan's cities plus Kabul; a northern and a southern Indian community, the former centered on Delhi and including Kathmandu and Thimphu and the latter including Mumbai, Kolkata, Hyderabad, Chennai, Bangalore; and two smaller communities for Dhaka and Chittagong in Bangladesh on the one hand and Male and Colombo on the other hand.

Betweenness centrality is clearly very concentrated: only 10 or so cities have a sizable value, making these into the key nodes for (re)producing the region's limited integration through infrastructure networks. Delhi, Mumbai and Lahore dominate, followed by Karachi, Chennai, Colombo and Dhaka.

Lahore has a relative fast road link with Amritsar. In addition, there is also a (low-frequency) transnational train service between Lahore and Amritsar/Delhi and a (low-frequency) flight between Lahore and Delhi. This makes the Lahore-Amritsar and Lahore-Delhi connections, although not very strong *per se*, into vital links for interconnecting Pakistan and India, and explains the large values for Delhi, Lahore and Amritsar.

Dhaka's gateway function is the consequence of being the only go-between for Chittagong's (and probably also most other cities in Bangladesh if the population threshold were to be lowered) connections with the rest of the network. This in turn fuels Kolkata's position, as this city – together with Delhi – functions as Dhaka's main gateway to the rest of the network.

And finally, Colombo functions as a gateway to Male given dense airline connections between both cities on the one hand, and Colombo's relatively strong air transport connections to the Indian subcontinent on the other hand. Given Colombo's solid connections with Chennai, the latter city also plays an important mediating role between Male/Colombo and the rest of the network.

Software (mandatory). The raw data is complied in the Excel CSV format and imported into the R statistical platform (R Core Team 2013) for processing. The data transformation, network analysis, and geographic visualization are subsequently performed in R using R-packages including *igraph* (Csardi and Nepusz, 2006) and *maps* (Brownrigg, 2013).

Acknowledgements

Map Design

A default rectangular projection in *maps* is chosen with equal longitudinal and latitudinal scales at the center of the map. The size of the nodes in the centrality map is determined as follows.

> 1. The maximum (normalised) centrality is denoted as max_centrality, the minimum positive (normalised) centrality is denoted as min_centrality. Note that the absolute minimum centrality will always be zero.
> 2. The city with min_centrality is drawn with a radius of 0.1 in user unit.
> 3. The node size of other cities is proportional to their centrality scores. More specifically, the node size for a city with a centrality of C is calculated with a radius of sqrt(C/min_centrality)*0.1
> 4. Because we want the cities with 0 centrality to be (somewhat) mapped, the node size for cities with zero centrality is calculated with a radius of sqrt((min_centrality/2)/min_centrality)*0.1. So even cities with no centrality can still be legible on the map.

References

Brownrigg, R. (2013). Maps: Draw Geographical Maps. Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka tpminka@media.mit.edu. available at http://CRAN.R-project.org/package=maps

Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. available at http://igraph.sf.net

Calderon, C. A., and L. Serven. 2004. The Effects of Infrastructure Development on Growth and Income Distribution. World Bank Policy Research Working Paper No. WPS 3400. Washington, DC: World Bank.

Clauset A., Newman, M.E.J. & Moore, C. (2004) Finding community structure in very large networks, Physical Review E, 70, 066111. DOI: 10.1103/PhysRevE.70.066111.

Hennemann, S., Rybski, D. & Liefner, I. (2012) The Myth of Global Science Collaboration. Journal of Informetrics 6(2): 217-225.

Hennemann, S. (2013) Information-rich visualisation of dense geographical networks. Journal of Maps, 9(1), 68-75.

Hennemann, S. & Derudder, B. (2013) An Alternative Approach to the Calculation and Analysis of Connectivity in the World City Network. Environment and Planning B, in press.

R Core Team. (2013) R: A Language and Environment for Statistical Computing. available at http://www.R-project.org

Sahoo, S., and K. K. Sexena. 1999. Infrastructure and Economic Development: Some Empirical Evidence. Indian Economic Journal 47(2): 54–66.

Shavitt. Y, & Shir, E. (2005) Dimes: Let the Internet measure itself. ACM SIGCOMM Computer Communication Review, 35: 71-74

Straub, S., C. Vellutini, and M. Walters. 2008. Infrastructure and Economic Growth in East Asia. World Bank Policy Research Working Paper No. 4589. Washington, DC: World Bank.

Taylor, P.J., Derudder, B., Hoyler, M., and Ni, P. (2013) New Regional Geographies of the World as Practised by Leading Advanced Producer Service Firms in 2010. Transactions, Institute of British Geographers 38(3): 497-511.

Tranos, E. & Nijkamp, P. (2013) The death of distance revisited: Cyber-place, physical and relational proximities. Journal of Regional Science, in press

Figure 1: The four network layers used in the construction of the composite network. Clockwise from the upper left corner: rail, road, air, and information technology. Edge width and darkness vary with the strength of the connection.

Figure 2: QAP correlation between the original network and the network after the application of a threshold.