Online Supplementary Material to Institutional Interconnections: understanding symbiotic relationships

NADIA VON JACOBI*

Said Business School, University of Oxford, Oxford, UK

Department of Political and Social Sciences, University of Pavia, Pavia, IT

This is an online appendix to the article *Institutional Interconnections: understanding symbiotic relationships*, published in the *Journal of Institutional Economics.* It includes more details on the sources of data and variables used in the analysis (Appendix A), a brief summary of how to replicate the construction of a correlation network (Appendix B), results of the sensitivity analysis that tested different significance thresholds for the correlation network (Appendix C), further details that relate to the investigation of correlation networks across different stages of municipal development (Appendix D) and their related network statistics (Appendix E), a section dedicated to quantile regressions, which includes introductory notes and the formal treatment of the analysis of asymmetric relationships (Appendix F). Appendix G overviews limits and potentials of the methodological strategy proposed in the paper.

*nadia.vonjacobi@sbs.ox.ac.uk or nadia.vonjacobi@unipv.it

1 APPENDIX A - THE MESO-LEVEL DATASET

The *meso-level dataset* contains information on economic structure, socio-demographic and institutional characteristics of Brazil's 5565 municipalities. The data set has been constructed on purpose to gather a significant amount of structural factors at the municipality level for Brazil. It combines publicly available measures of demographics, economic performance, labor market structure, institutional organization, productivity, local public finance and other municipal characteristics. All data refer to the year 2010, in some cases to 2009. The data sources that I use include the latest census available CENSO 2010 (IBGE), FAZENDADATA which comprise all public accounts of each municipality, the PERFIL DOS MUNICIPIOS (IBGE) - a municipality survey - which provides details on local institutions, their activity, organization and internal structure, and additional municipal information coming from IPEADATA on e.g. agricultural productivity, GDP per capita, export values and their recent growth rates. Last but not least, I include the municipality-level development indicators calculated by FIRJAN, which measure the advancements of all Brazilian municipalities in terms of employment, health and education.

In addition to the wealth of information provided by the original data sets, I have further developed a series of variables that comprise composite indicators, diversification indexes and network measures, in the attempt to grasp with greater detail how the local context is functioning. Aim of my additional variables is to capitalize the laudable effort that the Brazilian authorities have put into transparency and data availability. The accuracy and tidiness with which these data sets are put at disposal of the public allows to adopt an exploratory approach in which innovative measures of structural factors can be constructed.[1]

The *meso-level* data set for Brazilian municipalities represents an ideal informa-

[1]Constructing the *meso*-level data set has required months of work in terms of data collection and data management as the merge of the different sources has not always been easy due to differences in the shape of key variables.

tion resource for the implementation of correlation network analysis: not only because of abundance of variables and observations, but also because of the huge variability in municipal circumstances in Brazil. This data set therefore represents an excellent opportunity for the quantitative investigation of social structures and their interdependence.

Table 1 lists the 54 structural factors included in the analysis, including a brief description, the variable name included in some figures, the thematic area it may tentatively be assigned to, the number of observations, minimum and maximum values, and whether the variable has been log-transformed before computing correlation coefficients.

Table 1: Variables included in the analysis

NR	STRUCTURAL FACTOR	VARNAME	DESCRIPTION	AREA	OBS.	MIN	MAX	LOG
1	density population	popdensity	total residents/municipal area (in km^2)	D;G	5562	0.13	13030.5	x
2	dependency ratio	depratio	dependency ratio	D	5564	0.16	1.23	
3	nr of residents	poprestot	total resident population	D	5563	805	11253503	x
4	% of female residents	popfemper	female % on resident population	D	5563	30	54.2	
5	% of urban residents	popurbper	% of resident population in urban areas	D;H	5563	4.2	100	
6	agric share GDP	SHpibagro	share of agriculture in municipal GDP	ES	5564	0	0.81	
7	productivity perm cultivation	pprodlavperm	reais of product per hectar of permanent cult.	ES;G	5171	0	51.4	x
8	productivity temp cultivation	pprodlavtemp	reais of product per hectar of temporary cult.	ES;G	5439	0	179.1	x
9	diversification GDP	PIBmunDI	diversification of municipal GDP	ES	5564	0.18	0.73	
10	employment growth 10y	empgrowth0010	growth in employment rate 2000-2010 (%)	ES;H	5505	-37.8	98.2	
11	formality of economy	ESformdem2	ratio indirect taxes/GDP factor cost (no taxes)	ES	5560	0	0.78	x
12	industry share GDP	SHpibind	share of industry in municipal GDP	ES	5564	0.01	0.89	x
13	internal trade tax on tot taxes	SHICMStax	share of production and circulation (ICMS) tax	ES;G	5210	0	1	
14	labor-intensity formal economy	SHtrabIRRF	share of labor withholding tax (IRRF) on total IRRF	ES;INS	5192	0	1	
15	mun dependency on transfers	IdepTR2	net transfers (minus deductions) on municipal income	ES	5212	0.24	1.15	
16	% informal public employment	Iforpub	informal public workers on tot. public workforce	ES;INS	5548	0	0.975	x
17	prevalence permanent cultiv.	rtpermcult	ratio between permanent and temporary cultivation	ES;G	5171	0	1	
18	share public sector employment	rtadminWF	ratio of public employees on total workforce (15-59y)	ES;INS	5553	0.01	0.4	x
19	transfer inflows pcap	PCtransfcorr	per capita transfer inflow	ES;INS	5211	241	12935	x

G=Geography; H=History; D=Demography; TR=Tradition; ES=Econ. Structure; INS=Form.Institutions; HET=Soc.Heterogeneity; ATT=Soc.Attitude

Table 1: Variables included in the analysis

NR	STRUCTURAL FACTOR	VARNAME	DESCRIPTION	AREA	OBS.	MIN	MAX	LOG
20	change in inadequate sanitation	evHHsanINAD	rate of change 2000-10 in nr. of households	H	5154	-1	10	
21	density transportation services	km2transp	nr. of transport services/municipal area (in km^2)	G;INS	5564	0	0.83	x
22	petroleum royalties pcap	PCcotapet	per capita cota petroleo	G;ES	5211	0	351	x
23	territory resources	terres	indicator for resources on the territory	G	5565	0	0.969	
24	access to justice	Iaccessjus	composite index for access to justice	INS;TR	4802	0	0.83	
25	admin costs on mun spending	SPshadmin	share of admin costs on total mun spending	INS	5212	0	1	
26	diversification public income	RECmunDI	diversification of public income sources	INS;ES	5211	0.01	0.76	x
27	diversification transfers	TRmunDI	diversification of transfer sources	INS	5212	0	0.74	
28	institutional collaborations	P1000IAIsum	nr. of, across all themes (p1000inh)	INS	5563	0	8.9	x
29	leakage earmarked res education	SPlkedu	per cap. earmarked resources-per cap. spending	INS	5593	0	10	x
30	nr activities local police	nratpolicia corr	nr. of activities of the local police	INS	864	1	22	
31	nr transfer contacts	NRinstcont	nr of institutional contacts through transfer-flows	INS	5212	28	55	
32	participatory councils allthemes	ICMsum	presence and strength of, all themes	INS	5565	0	58	
33	participatory councils p1000inh	P1000ICMsum	density of, (p1000inh)	INS	5563	0	20	x
34	share taxes on mun income	SHtaxRC	share of taxes on municipal current income	INS; ES	5212	0	0.69	x
35	spending health pcap	SPpcsaude	per capita spending on health	INS	5211	0	1928	
36	spending housing share	SPshhou	share of spending on housing on total mun spending	INS	5212	0	0.71	x
37	spending industrialpolicy share	SPshindpol	industrial policy costs on total mun spending	INS; ES	5212	0	0.4	x
38	spending publicgoods share	SPshbenspub	spending on pub. goods on mun spending	INS	5212	0	0.63	
39	transfers from FedUnion	SHU	share of transfers coming from the Federal Union	INS	5212	0	1	

G=Geography; H=History; D=Demography; TR=Tradition; ES=Econ. Structure; INS=Form.Institutions; HET=Soc.Heterogeneity; ATT=Soc.Attitude

Table 1: Variables included in the analysis

NR	STRUCTURAL FACTOR	VARNAME	DESCRIPTION	AREA	OBS.	MIN	MAX	LOG
40	transfers from State	SHE	share of transfers coming from the State	INS	5212	0	0.86	
41	age of mayor	Iyouthpref	age of mayor (normalized)	ATT; TR	5546	0	1	
42	education of mayor	eduprefeito	educational level of the mayor (normalized)	ATT; H	5564	0	1	
43	female empowerment	femempower	female empowerment (normalized index)	ATT; TR	5564	0.11	0.87	
44	nr art groups	PCIgroupart	per 1000 inh. nr. of art groups	ATT; TR	5563	0	5	x
45	nr cultural equipment	Iequipcult	nr. of, available in the municipality	ATT	5565	0	17	
46	% deaths homicide	propmorthom	% of deaths due to homicide (2009)	ATT; D	3339	0	0.67	
47	% deaths suicide	propmortsuic	% of deaths due to suicide (2009)	ATT; D	2507	0	0.5	x
48	political competition for mayor	candvot	nr. of candidates voted for mayor	ATT	5545	1	12	
49	child poverty	childpov32010	%, living with analph. adult & inadeq. sanitation	HET	5563	0	58.1	x
50	ethno-age fractionalization	comNTWKn	likeliness of missing comunication	HET	5565	0	1	
51	% adult illiterates	totillper	share of illiterates aged 10 or older	HET	5052	0.2	49.5	x
52	% single headed households	resphhper	% of households with only one adult responsible	HET	5564	25.2	100	
53	population diversity index	popHETnorm	age and family type diversity (norm)	HET	5473	0	1	
54	wage diversity index	popDIwagen	sum of wagegaps across ethnic groups (norm)	HET	5565	0	1	x

G=Geography; H=History; D=Demography; TR=Tradition; ES=Econ. Structure; INS=Form.Institutions; HET=Soc.Heterogeneity; ATT=Soc.Attitude

2 APPENDIX B - HOW TO CONSTRUCT A CORRELATION NETWORK DATASET

The construction of a correlation network can be subdivided into three main phases, two of data preparation and one of analysis. First build a dedicated dataset, which collects institutional and other structural variables at a given level of analysis, here municipalities in Brazil (phase 1).

In phase 2, calculate pairwise correlations among all institutional and structural variables included in the analysis in order to construct a relational dataset (so-called *edgelist*) to which network statistics can be applied. The unit of analysis of an edgelist is the single relation, in this case the relation (*edge*) between two structural factors (*nodes*) is the correlation coefficient that can be detected among the available observations (e.g. in my Brazilian dataset, computed on the sample of 5565 municipalities).[2] Where data abundance is given, I prefer using weighted correlation networks, which allow for the *quantification* of the interconnection, namely the absolute value of the correlation coefficient. If qualitative or more reduced samples were to be used, it is possible to opt for *unweighted* correlation networks, which imply thresholding the relation to some predetermined value.

Phase 3 requires the reshaping of the dataset in the following way: eliminate the original variables and transform columns reporting correlation coefficients into observations (rows). At this point it is possible to apply network statistics to the obtained edgelist. A series of different software packages is today able to compute standard network statistics. I have made use of the package *netsis* (Miura, 2012) designed for STATA, which allows for the computation of network statistics on weighted networks. Note however that for its application to correlation networks, it is necessary to use the inverse of the correlation

[2]It is possible to apply restrictions, e.g. to the most significant cases, such as correlation coefficients that are statistically significant at the 5% level. In such case a Pearson test can be used for detection. In this analysis, only correlation coefficients with statistical significance at the 1% level have been included in the correlation network.

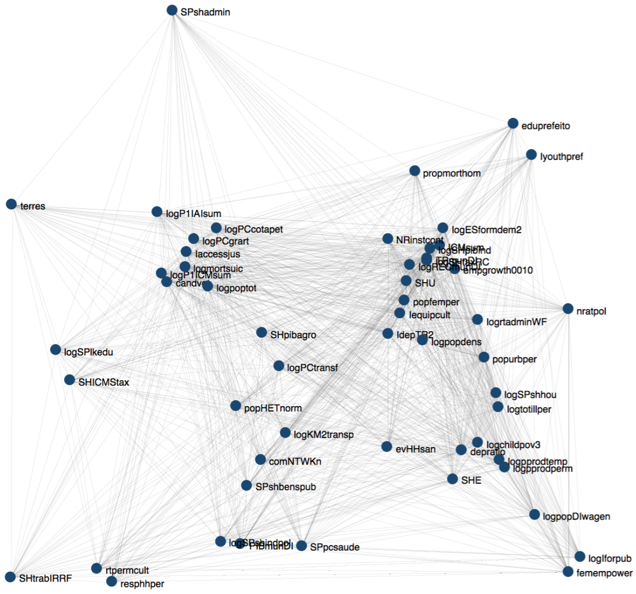


Figure 1: Complete correlation network, with correlations at statistical significance 1%

coefficient in all network statistics that make use of the concept of *paths*, such as the closeness or betweenness degree. The plots included in this article have been designed using the new set of STATA commands designed for social network analysis, *nwcommands* (Grund et al., 2015). Figure 1 displays the complete correlation network among municipal structural factors considering only correlations with statistical significance at 1%.

3 APPENDIX C - SENSITIVITY ANALYSIS ON STATISTICAL SIGNIFICANCE OF CORRELATION COEFFICIENTS

This appendix briefly reports on a sensitivity analysis that has compared findings in terms of highest ranking structural factors according to some network statistics when using two different specifications of the correlation network. In the first case, all correlation coefficients with statistical significance of up to 5% have been included. The second specification is more restrictive as it only includes correlation coefficients with statistical significance at 1%. As can be seen in figure 2, the kind of factors that rank among the highest 20 of two principal centrality measures tend to be the same. The measure of unweighted degree centrality displays more re-rankings in relative terms (factors included in the table but changing their overall rank within it). The lower panel (2b) shows that weighted degree centrality, which is a more reliable measure as it also considers the strength of correlation, is basically unaffected. Unweighted degree centrality, on the other hand, which only counts the number of connections is naturally more affected by a threshold that reduces the number of edges considered (in this case falling from 2478 to 2368). In figure 2, structural factors are highlighted in the colours that correspond to their thematic area introduced in the main text and reported in table 1. In line with the findings of the sensitivity analysis, all computations based on correlation networks reported in the main text use the cut off at 1% statistical significance.

Unweighted Degree Centrality

5% significance			1% significance		
structural factor	rank	value	structural factor	rank	value
formality_of_economy	1	0.981	transfer_inflows_pcap	1	0.962
mun_dependency_on_transfers	2	0.962	mun_dependency_on_transfers	1	0.962
transfer_inflows_pcap	2	0.962	agric_share_GDP	1	0.962
agric_share_GDP	2	0.962	transfers_from_FedUnion	2	0.943
diversification_public_income	2	0.962	share_taxes_on_mun_income	2	0.943
share_taxes_on_mun_income	2	0.962	diversification_public_income	2	0.943
nr_of_residents	2	0.962	nr_cultural_equipment	2	0.943
%_of_urban_residents	2	0.962	ethno-age_fractionalization	3	0.925
nr_transfer_contacts	2	0.962	formality_of_economy	3	0.925
diversification_transfers	3	0.943	share_public_sector_employment	3	0.925
transfers_from_FedUnion	3	0.943	%_of_urban_residents	3	0.925
nr_cultural_equipment	3	0.943	spending_publicgoods_share	3	0.925
ethno-age_fractionalization	3	0.943	nr_of_residents	3	0.925
share_public_sector_employment	4	0.925	nr_transfer_contacts	4	0.906
spending_publicgoods_share	4	0.925	participatory_councils_allthemes	4	0.906
%_deaths_suicide	4	0.925	diversification_transfers	4	0.906
diversification_GDP	4	0.925	**density_population**	4	0.906
industry_share_GDP	4	0.925	%_deaths_suicide	4	0.906
participatory_councils_allthemes	4	0.925	dependency_ratio	4	0.906
nr_art_groups	4	0.925	petroleum_royalties_pcap	5	0.887
TOTAL NR. OF SIGNF. CONNECTIONS	**2478**			**2368**	

(a) Unweighted Degree Centrality, ranks using correlations at statistical significance of 5% (left) and of 1% (right)

Weighted Degree Centrality

5% significance			1% significance		
nr_of_residents	1	0.33	nr_of_residents	1	0.328
diversification_public_income	2	0.304	diversification_public_income	2	0.303
share_taxes_on_mun_income	3	0.3	share_taxes_on_mun_income	3	0.3
nr_cultural_equipment	4	0.298	nr_cultural_equipment	4	0.298
formality_of_economy	5	0.29	formality_of_economy	5	0.288
mun_dependency_on_transfers	6	0.283	mun_dependency_on_transfers	6	0.283
petroleum_royalties_pcap	7	0.283	petroleum_royalties_pcap	7	0.282
%_deaths_suicide	8	0.267	child_poverty	8	0.267
child_poverty	9	0.267	transfers_from_FedUnion	9	0.267
transfers_from_FedUnion	10	0.267	%_deaths_suicide	10	0.266
participatory_councils_p1000inh	11	0.265	participatory_councils_p1000inh	11	0.265
share_public_sector_employment	12	0.265	share_public_sector_employment	12	0.265
%_of_urban_residents	13	0.264	%_of_urban_residents	13	0.263
transfers_from_State	14	0.259	transfers_from_State	14	0.259
institutional_collaborations	15	0.243	institutional_collaborations	15	0.241
density_population	16	0.24	density_population	16	0.239
nr_art_groups	17	0.239	dependency_ratio	17	0.239
dependency_ratio	18	0.239	transfer_inflows_pcap	18	0.239
transfer_inflows_pcap	19	0.239	nr_art_groups	19	0.238
agric_share_GDP	20	0.233	agric_share_GDP	20	0.233
TOTAL NR. OF SIGNF. CONNECTIONS	2478			2368	

(b) Weighted Degree Centrality, ranks using correlations at statistical significance of 5% (left) and of 1% (right)

Figure 2: Results of the sensitivity analysis in which factors ranking highest in terms of unweighted (upper) and weighted (lower) degree centrality are compared across two different specifications of the correlation network

4 APPENDIX D - CORRELATION NETWORKS AT DIFFERENT STAGES OF DEVELOPMENT

In what follows, I provide details on the analysis that investigates how the identification of correlation networks and their centroids varies at different levels of development of Brazilian municipalities. The employment pillar of the FIRJAN index of municipal development has been used to subdivide the population of Brazilian municipalities into sub-groups along the five quintiles of the chosen development indicator. Figure 3 shows the distribution of municipal values of the employment pillar in comparison to those capturing corresponding development levels in education and health: these latter ones are not significantly different across the chosen development quintiles - hinting for strong national convergence in these sectors (von Jacobi (2014); WB (2016)). Yet, a slight trend for higher performance in health and education can be detected in association with greater income and formal employment.

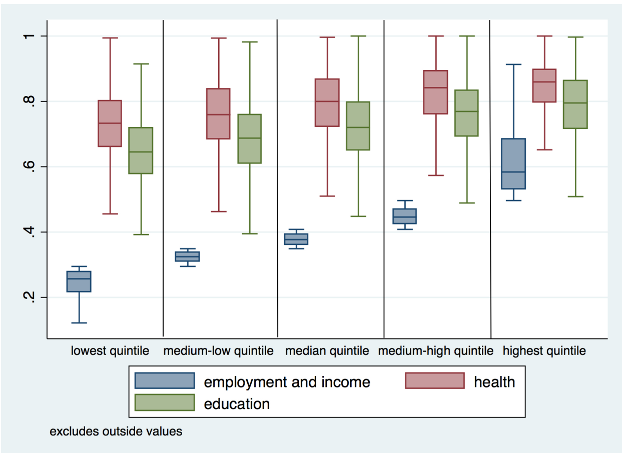


Figure 3: Distribution of municipal levels in human development measures, by quintiles of economic development, excluding outliers *Source*: Elaboration by the author based on FIRJAN, 2009

The following figures display the network graphs computed at each of the five different levels of development of Brazilian municipalities. For a legend that helps identifying the

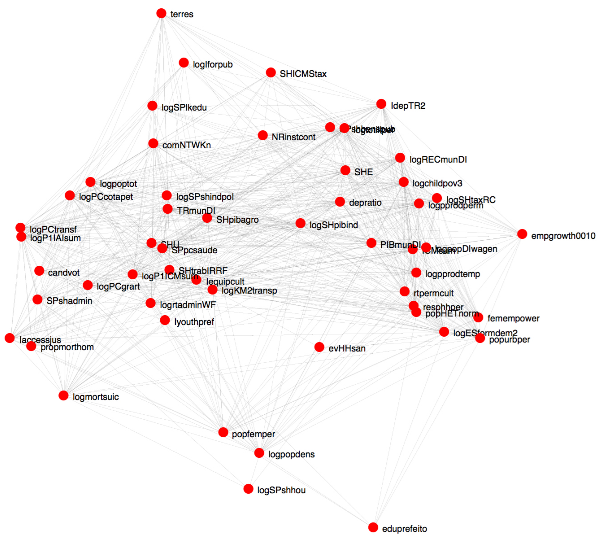
Table 2: Network characteristics at different levels of development

development quintile	nr. of edges	density of network	degree centralization	betweenness centralization
lowest	724	0.525	0.273	0.018
medium-low	765	0.535	0.287	0.018
median	812	0.589	0.287	0.037
medium-high	904	0.632	0.245	0.017
highest	943	0.659	0.237	0.012

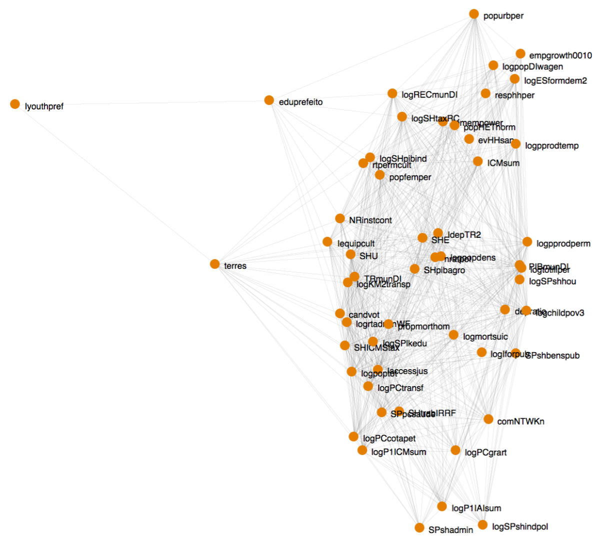
Source: Author's elaboration

meaning of labels, please see table 1. All five plots of the correlation networks show that the density is rather high.

Table 2 summarizes some key data on the structures of the correlation networks computed on the five sub-samples of municipalities. As mentioned in the main text, the five networks are basically equivalent, apart from an observable increase in density and decrease in degree centralization at higher levels of development. In figures 5b and 6 it is possible to see greater network centralization, which results in more factors being located at central positions - automatically implying decreased degree centralization of single factors.

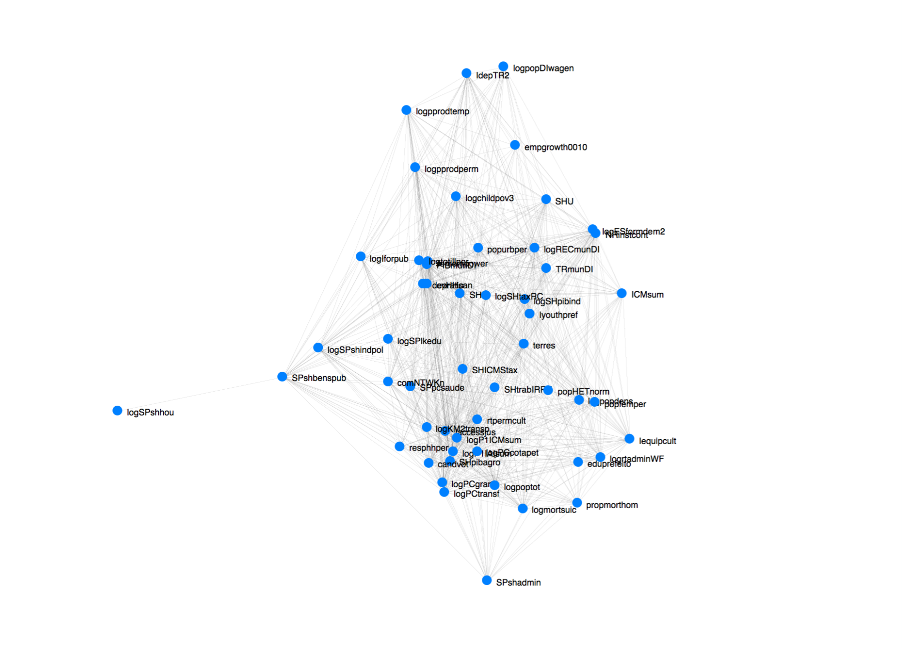


(a) Complete correlation network computed on municipalities at the lowest quintile of development

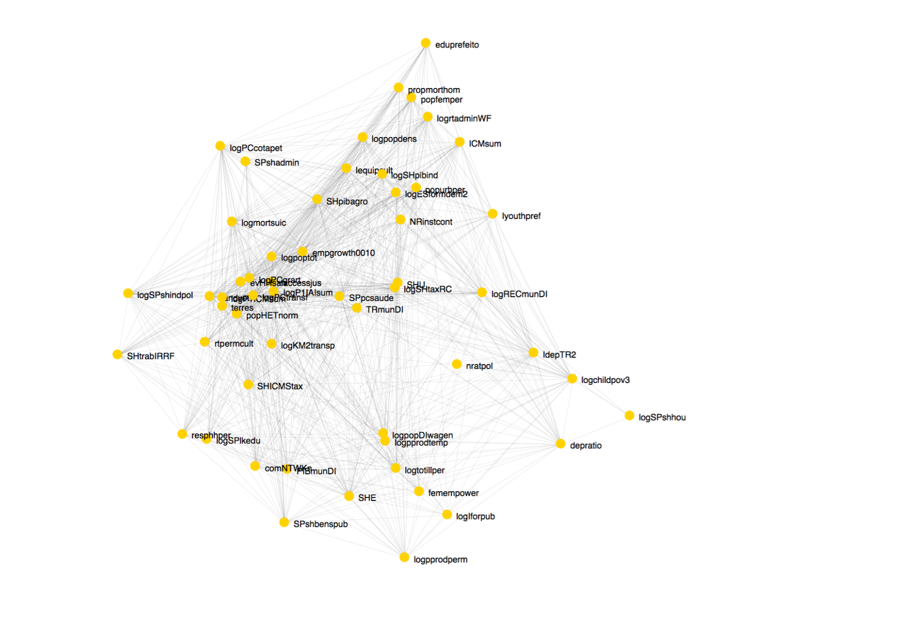


(b) Complete correlation network computed on municipalities at the medium-lower quintile of development

Figure 4: Complete correlation networks computed on sub-groups of Brazilian municipalities, according to their level of economic development



(a) Complete correlation network computed on municipalities at the median quintile of development



(b) Complete correlation network computed on municipalities at the medium-higher quintile of development

Figure 5: Complete correlation networks computed on sub-groups of Brazilian municipalities, according to their level of economic development (II)

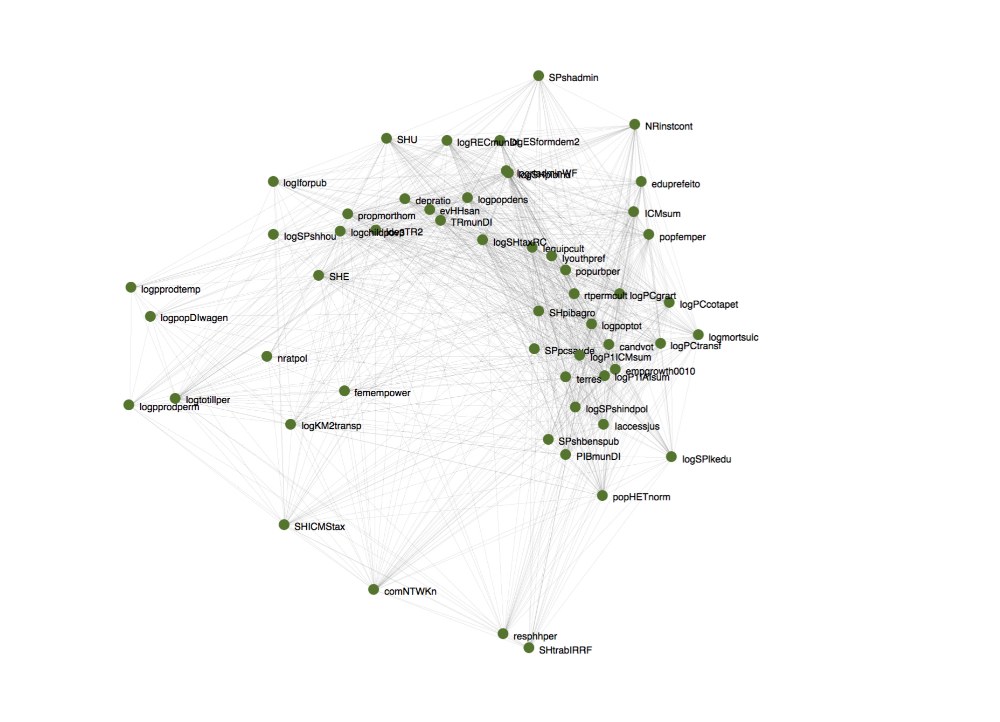


Figure 6: Complete correlation network computed on municipalities at the highest quintile of development

5 APPENDIX E - CENTROIDS ACROSS DEVELOPMENT STAGES

Table 3 below reports the ranks from 1 to 20 that structural factors assume in the computations of four network statistics (unweighted and weighted, closeness and betweenness centrality). These computations are run on five subsamples of the overall population of Brazilian municipalities: the five sub-samples correspond to the quintiles of the overall distribution in terms of a proxy for economic development (see main text and Appendix D). In table 3, it is possible to track how ranks of structural factors - within the same and across different network statistics - change for municipalities with different levels of development. Where ranks are missing in the table, this implies that they are not among the first 20. Lower ranks have not been reported to facilitate the identification of factors that persist at top ranks across different development stages. In case of the ranks of unweighted degree centrality, different factors display the same value of the network statistics, meaning they have an identical amount of significant correlations to other factors in the network. A detailed description of results is included in the main text.

Table 3: Centroids across different development levels

	Unweighted D					Weighted D					Closeness D					Betweenness D				
Structural Factor	D1	D2	D3	D4	D5	D1	D2	D3	D4	D5	D1	D2	D3	D4	D5	D1	D2	D3	D4	D5
nr cultural equipment	1	6	9	4	1	12	15	12	7	4	16	14	11	10	6	15	4	8	14	3
% adult illiterates	2	5	6	5		6	8	7	19		4	9	7	18		12	18	10	13	8
nr of residents	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4
transfers from State	3	4	3	7		5	6	6	20		5	3	6			4	3	3	5	13
participatory councils p1000inh	4	5	2	3	1	3	4	2	2	2	2	5	5	3	2	3		15	18	15
dependency ratio	4	3	6		5	11	11	13			13	12	9		19	10	7	16		
child poverty	5	2	5		4	7	5	5	18	16	6	6	4	13	12	5		2	3	1
mun dependency on transfers	5	4			1	10	12	19		12	12	16	18		13	18			17	9
diversification public income	5	7	7	3	2	9	13	15	11	5	11	15	16	12	7	8		9	8	19
density transportation services	6	8				19										16			12	12
petroleum royalties pcap	6	5	4	7	6	2	3	4	4	3	3	7	8	2	3	6	13	11	6	6
ethno-age fractionalization	6	6	8	7		14	9	9			8	4	2	7			6	7	4	14
transfers from FedUnion	7	9		8	6	13	17		15		7	10	10	19		11	10	18		10
formality of economy	8		6	5	6			18	12	18	18	17	17	11	18		20		19	
industry share GDP	8			7																
spending health pcap	8	8				8	7	11	17		10	8	13							
share taxes on mun income	9	7	5	1	3	20	19	16	6	10			20	8	8	14	9	20	9	5
agric share GDP	9	6	5	2	1			17	5	8		18		6	9	17	8	13	7	11
diversification transfers	9	9				16	18				15	13	15			7		17	11	
% of urban residents	9		9	6	3				13	7	20			5	5	13	5	4	2	17
population diversity index				7													15	19		20
transfer inflows pcap		3	3	2		4	2	3	3	15	9	2	3	4	17		12	5	15	7
institutional collaborations			4	1	3	17	14	8	8	6	19	19	14	15	11					
density population					4					11				20	4	19	16	12	10	2
% deaths suicide			9	7		15	10	10	9	14	14	11	12	9	15	2	17		20	16
nr art groups			9	5	6		20	14	10	9	17	20	19	14	10					
share public sector employment					6				16	13				16	14					18
diversification GDP		5	6				16													
% of female residents					6					17					16	9	19	14		
access to justice		6	5	3	6			20	14	19				17	20					
participatory councils allthemes					6					20										
spending publicgoods share																20		6	16	
territory resources																	11			
education of mayor																	14			

6 APPENDIX F - QUANTILE REGRESSION LOOPS TO BUILD A DIRECTED NETWORK

This appendix provides some synthetic description of how quantile regressions work. It is meant to be of use for those who are unfamiliar with the concept. It then includes details on the formal treatment elaborated to identify asymmetric relationships in correlation networks.

6.1 A brief introduction to Quantile Regressions

Quantile regressions can broadly be understood as follows: "what the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of xs. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily, this is not done, and so regressions often give a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions" (Mosteller and Tukey, 1977 in (Koenker, 2005, p.3)).

Quantile regressions allow to focus on noncentral locations on the response distribution. The *quantile* is to be understood as a generalizing term for the more specific *quartiles, quintiles, deciles and percentiles*: "the pth quantile denotes that value of the response below which the proportion of the population is p" (Hao and Naiman, 2007, p.3).[3] By not restricting the analysis to the conditional mean, quantile regressions are a more robust technique for variables that are not normally distributed, as often happens in socio-economic and institutional analysis. They remain an extension to the linear regression model, though.

[3]This is in line with a cumulative density funcion F_y that for each value of y provides us with the proportion of the population for which $Y \leq y$ (Hao and Naiman, 2007, p.7).

6.2 Two-way quantile regression loops

To identify symbiotic relationships with asymmetric character, it is necessary to construct a *directed* network, meaning that it is possible to identify directions of influence between different nodes. This requires refinement of the computation of the interconnection network, in particular the construction of a network based on regression models (Horvath, 2011, ch.13). To properly capture asymmetric symbiotic relationships, I make use of two-way quantile regressions. I compute the following quantile regression for each variable included in table 1:

$$y_i = \beta_0^{(p)} + \beta_1^{(p)} x_i + \sum_{j=1}^{4} \delta_j^{(p)} DEV_{ij} + \epsilon_i^{(p)} \tag{1}$$

where the quantiles p are set at five different values of the distribution of the y variable: *p20; p35; p50; p65; p80* and y_i and x_i represent any structural variable of table 1 for municipality i. In order to reduce the potential effect of omitted variables in the estimation of the direct relations, I include the quintile of development to which the municipality belongs (as introduced in appendix D) as control factor.Every single relation could benefit from an own specification of the regression model. However, different specifications would make the comparison across relations more difficult, which is exactly the intrinsic goal of a correlation network: it implies a more systemic view on the totality of relations and therefore needs to treat them in a way that makes them equivalent, to some extent - although that necessarily also implies greater superficiality on the details of single relations. A second argument against specifying each relation singularly relates to computational costs: given the high amount of regressions to specify $((54 * 53) = 2862)$, this would be too time intensive indeed. For each y, I estimate five quantile regression models for which the pth conditional *quantile* given x_i is

$$Q^{(p)}(y_i|x_i) = \beta_0^{(p)} + \beta_1^{(p)} x_i + \sum_{j=1}^{4} \delta_j^{(p)} DEV_{ij} + \epsilon_i^{(p)} \tag{2}$$

where the pth quantile of the error term is zero.[4] To assure significance of estimations, I only preserve coefficients with a p-value of $p <= 0.05$. By looping through the entire list of variables of table 1, I compute two quantile regressions for each possible pair of variables - two because the dependent and the independent variables are switched for each pair.

6.3 Commensalist Structural Factors

After running the quantile regression loops, the final part of the analysis investigates the asymmetric character of the pairwise relations. For simplicity reasons, I focus on commensalist relations only. This implies comparing the quantile regression estimations of the regression in which y is the dependent factor with those of the regression in which x is the dependent factor. Let's imagine a situation in which $A = x$ and $B = y$, where y is the dependent variable. Five of the estimated coefficients of equation 2 are of particular interest, namely

$$\beta_{1A}^{(p20)}; \beta_{1A}^{(p35)}; \beta_{1A}^{(p50)}; \beta_{1A}^{(p65)}; \beta_{1A}^{(p80)} \tag{3}$$

where each coefficient describes how factor A explains variability in the dependent variable B at a specific moment of its distribution. The subscript A merely serves for clarity in assigning the estimated coefficient to a particular explanatory factor, in this case A. The quantiles $p20, p35, p50, p65, p80$ refer to the distribution of factor B. If

$$\beta_{1A}^{(p20)} < \beta_{1A}^{(p35)} < \beta_{1A}^{(p50)} < \beta_{1A}^{(p65)} < \beta_{1A}^{(p80)} \tag{4}$$

we are observing a situation in which at higher levels of B (higher moments of its distribution), A is more and more relevant. Somehow, at higher levels of B, B is more dependent on A for its growth. I now invert the dependent and independent variable in

[4] In line with (Hao and Naiman, 2007, p.29)). Error terms at different quantiles are not neccesarily i.i.d.

order to have $B = x$ and $A = y$. If at higher levels of A, A is not more dependent on B for its growth, then we observe an asymmetric relationship in which B is a *commensalist.* The following conditions are derived in order to identify commensalism. Commensalism of B over A is detected if:

- the correlation coefficient between A ad B is positive
- $\beta_{1A}^{(p80)} > \beta_{1A}^{(p20)}$
- $\beta_{1B}^{(p80)} \geq \beta_{1B}^{(p20)}$
- $\{\beta_{1A}^{(p80)} - \beta_{1A}^{(p20)}\} > \{\beta_{1B}^{(p80)} - \beta_{1B}^{(p20)}\}$

Figure 7 shows all commensalist relations that can be detected among structural factors at the municipality level in Brazil. This directed network is far less dense than the correlation network computed initially (see figure 1). Arrows stand for the direction of "benefit", going towards the factor gaining more from the asymmetric relationship. The reduced density of the directed network may in part be due to the fact that estimations of commensalist relations are highly conservative: only converging estimates and relations in which all ten computed coefficients were significant at the 5% level have been included. Including a control factor reduces the amount of converging estimations due to the limitation of sub-samples for each of the estimated quintiles.

More detailed network graphs on the two major components of the directed network are included in the main text. Table 4 reports each factor involved in a commensalist relationship and the number of cases in which it assumes the role of the beneficiary *versus* the role of the benefactor.

Commensalist	beneficiary	benefactor
Institutional Collaborations	4	0
Share of Admin Costs	3	0
Spending on Public Goods	3	0
Inflow of Transfers	3	5
Share of Public Sector Employment	3	2
Spending on Health	2	5
Transfers from State	2	1
Share of Taxes on Mun Income	2	1
Incidence of Art Groups	2	1
Diversification of municipal GDP	1	0
Share of Circulation Tax on Total	1	2
Transfers from Federal Union	1	1
Dependency Ratio	1	1
Diversification of Public Income	1	0
Agricultural Productivity	1	0
Share of Adult Illiterates	1	0
Spending Leakage in Education	0	1
Agricultural Share in GDP	0	3
Population Density	0	1
Ratio permanent to temporary cultivation	0	1
Density Transportation Services	0	1
Ethno-age Fractionalization	0	1
Informality in Public Employment	0	1

Table 4: Factors engaged in commensalist relationships, total nr. of situations in which the factor is receiving or providing benefit to another factor

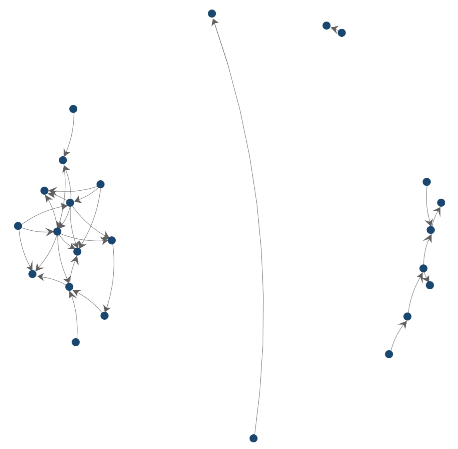


Figure 7: Complete network of commensalist relationships

6.4 Comparing asymmetric symbiotic relationships with supermodularity conditions

As mentioned in the main text, the investigation of symbiotic relationships has important commonalities with those studies of the institutional complementarities literature that adopt a co-evolutionary perspective on them (e.g. Aoki (2001), Battistini and Pagano (2008)). To better place the contribution of the study within this literature, this section puts the formal treatment so far presented into relation with standard supermodularity conditions, as treated in (Aoki, 2001, p.226).

Supermodularity conditions serve as overarching framework for games in which individual choices in one particular domain (such as e.g. the market) are dependent on parameters defined in another domain (e.g. public policy). One domain resembles the *institutional environment* within which individual maximization of payoffs has to occur. The interdependence outlined by Aoki is highly compatible with the one envisaged by symbiosis, in which some sort of interdependence - without specification of functionality - is present. Aoki introduces the supermodularity conditions that frame complementarity by drawing on Topkins (1978, 1998) and Milgrom and Roberts (1990), all in (Aoki, 2001, p.226). In two different domains, D and G, different agents M and N, respectively make choices that lead to the institutionalization of an endogenous rule. All agents are hypothesized to choose between two different rules, Σ^* or Σ^{**} (in the case of D) and Λ^* or Λ^{**} (in the case of G).

Payoff functions (u in D and v in G) are assumed to be identical within each domain in this standard specification and the following conditions are derived:

$u(\Sigma^*; \Lambda^*) - u(\Sigma^{**}; \Lambda^*) \geq u(\Sigma^*; \Lambda^{**}) - u(\Sigma^{**}; \Lambda^{**})$

$v(\Lambda^{**}; \Sigma^{**}) - v(\Lambda^*; \Sigma^{**}) \geq v(\Lambda^{**}; \Sigma^*) - (\Lambda^*; \Sigma^*)$

These conditions that underpin the complementarity of the two domains enshrine that, for each agent in domain D, Σ^* is the more convenient choice of rule in any case in which the rule Λ^* predominates. The existence of Λ^* constitutes an exogenous factor that favours the development of the endogeneous rule Σ^*. Similarly, for each agent in domain G, Λ^{**} is the more convenient choice of rule in any case in which the rule Σ^{**} prevails. In this case, it is Σ^{**} that resembles an exogenous factor that favours the development of the endogeneous rule Λ^{**}.

Supermodular conditions can also be used to express symbiotic relationships, as outlined in what follows: consider again a situation in which B is commensalist on A (as in section 6.3). The two social structures can be chosen both, by the agents belonging to two different domains, respectively identified by the choice paramenters σ and ϕ. The binary options for B and A are therefore to be present or absent.

Commensalism of B over A is present if the following two conditions hold:

$u(\sigma B; \phi A) - u(\sigma A; \phi A) > u(\sigma B; \phi B) - u(\sigma A; \phi B)$

$u(\phi A; \sigma B) - u(\phi B; \sigma B) = u(\phi A; \sigma A) - u(\phi B; \sigma A)$

In presence of A, the utility of B is greater. However, the same does not hold reversely: the presence of B does not constitute an incentive for choosing A. If we opt for representing the binary options for B as present B^+ or absent B^0, and for A, respectively as A^+ and A^0, we obtain the following conditions:

$$u(B^+; A^+) - u(B^0; A^+) > u(B^+; A^0) - u(B^0; A^0)$$
$$u(A^+; B^+) - u(A^0; B^+) = u(A^+; B^0) - u(A^0; B^0)$$

In presence of A, the utility of B is greater than if A is not present. However, the same does not hold reversely: the presence of B does not constitute an incentive for choosing A. Therefore, B is commensalist (and dependent) on A, but A is neutral/independent from B.

With reference to section 6.3, it should be noted that the formal treatment used there refers to coefficients estimated in quantile regression. It therefore adopts a data-driven approach. Theoretical treatment such as derived from Aoki's work, on the contrary help envisaging the underlying mechanism that is understood by symbiosis, in this case commensalism.

Note that Aoki's supermodularity conditions are concerned with "the property of incremental payoffs with respect to a change in parameter value" (Aoki, 2001, p.226). They are therefore adequate to investigate drivers of commensalism that relate to human choices, but may not be able to explain situations of commensalisms that are themselves dependent on exogenous, third factors such as environmental conditions. Their advantage lies in constituting a micro framework for explaining why specific commensalist relations may appear. Such framework can also be used to express parasitism (not treated here).

7 APENDIX G - LIMITS AND POTENTIALS OF THE METHODOLOGY

This final appendix discusses the advantages and disadvantages of using correlation networks to study institutional interconnections. Adopting a network perspective departs from the assumption that all structures characterizing a context are interdependent in possibly complicated ways: this provides fertile ground for an investigation of institutional interconnections that is informed by complexity. The use of correlation networks should be viewed as a data reduction technique that allows for the identification of centroids, which most likely play a major role for a series of institutional and structural factors of a given context. It can also serve for grasping the degree of connectivity that binds different social structures together. The implementation of two-way quantile regression loops represents an extension that seems able to identify asymmetric relationships, through which it is possible to construct *directed* networks, informing on the underlying structure of existing relations. As (Horvath, 2011, p.4) puts it, it is natural to use network methods when one tries to model pathways - and pathways open up new doors for the investigation of dynamics.

Some limitations of the methodology need to be highlighted, however: it requires abundance of quantitative information, not always or commonly available in institutional analysis. One possible strategy is to scale down the unit of analysis from the country to the subnational level, with obvious implications for the kind of structural factors - particularly institutions - that can be included. The approach may help identify *lacunae* in data collection. Further, a subnational application allows to focus on the study of components of institutional arrangements, which could help disentangle the "composite" nature of institutions, which according to Sindzingre (2014) is an underexplored area within institutional analysis.

Care is necessary also in the selection of variables. While cutting the correlation network at different levels of significance does not substantially alter results, a more delicate issue is the inclusion of factors. This requires a delicate calibration between a significant amount of variables and the avoidance of redundancy of measured entities, as this may affect the identification of centroids. A possible strategy to deal with this issue is to use sub-networks among thematically close variables in order to identify so-called *hubs*.[5] Previous knowledge of potentially relevant factors can inform and simplify such

[5] Here, a correlation sub-network has been run on public spending variables, selecting the most relevant

choice, but for exploratory analysis it may also be interesting to include less-studied factors, to detect unexpected connections. The final selection of factors therefore clearly depends on the specific research interests. The exploratory nature of the methodology mainly sheds light on *complexity*: this is very adequate when used to explore (and to select) potentially relevant links for subsequent in-depth analysis.

The extension to quantile regressions should also be used with care: direct causal relationships - despite of an existing asymmetry - may be tainted by possible statistical noise in single two-way relations. While the inclusion of control factors and the restriction to highly significant coefficients has proven to be possible, this also reduces the amount of asymmetric relationships that can be found. Again, a delicate calibration is necessary.

Apart from these limitations, the methodology represents a first step for a series of potential future elaborations. Many of these can build upon continuous advancements within the field of gene co-expression analysis, such as the identification of subclusters, so called modules, which can be interpreted as sub-systems of interconnections. Further, more investigations could go into the context-dependency of the network structures, e.g. investigating which centroids gain or lose importance at different degrees of urbanization or according to other features deemed relevant. Other extentions may be inspired from social network analysis, which is continuously evolving in conceptual and computational terms. The analysis of so-called *structural holes*[6] for example represents an interesting area of study, as nodes next to such holes may be important leverage points for policy. More investigations could also go into the explanation of the network structure: dynamic and inferential network analysis is likely to become more relevant in the future. Greater details on dynamics may improve our understandings of how interconnections evolve and why this may be related to development. The conceptual work on symbiotic relationships presented here may provide inspiration for the modelling of network dynamics. An

hub, namely spending on health, and then other variables that were least correlated to this hub.

[6]Studied e.g. by Ahuja, 2000; Burt, 1992, 2004; Podolny and Baron, 1997, in Green Jr and Wasserman (2013).

empirical investigation of parasitic relationships may be the next step in this direction.

Correlation networks have an intrinsic advantage over more standard network analysis: they allow for comparison across different networks. In gene co-expression analysis, such comparison is applied to different contexts - notably different organic tissues, such as liver vs. brain. For a wider application within the social sciences, the application of the technique could allow to compare findings among similar/identical measures across different socio-cultural contexts.

References

Aoki, M. (2001). *Toward a comparative institutional analysis.* MIT press.

Battistini, A. and Pagano, U. (2008). Primates' fertilization systems and the evolution of the human brain. *Journal of Bioeconomics*, 10:1–21.

Green Jr, H. D. and Wasserman, S. (2013). Network analysis: a definitional guide to important concepts. *Oxford Handbook of Quantitative Methods in Psychology, Vol.1.*

Grund, T. et al. (2015). Social network analysis using stata. In *United Kingdom Stata Users' Group Meetings 2015*, number 21. Stata Users Group.

Hao, L. and Naiman, D. (2007). *Quantile Regression.* Sage Publications, Thousand Oaks, California.

Horvath, S. (2011). *Weighted network analysis: applications in genomics and systems biology.* Springer Science & Business Media.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, New York.

Miura, H. (2012). Stata graph library for network analysis. *Stata Journal*, 12(1):94–129.

Sindzingre, A. N. (2014). Institutions as a composite concept: Explaining their indeterminate relationship with economic development. *Greenwich, World Interdisciplinary Network for Institutional Research Inaugural Conference, 11-14 September.*

von Jacobi, N. (2014). Can the context mediate macro-policy outcomes? Technical report, UNU-WIDER, Helsinki, Finland.

WB (2016). Retaking the path to inclusion, growth and sustainability. Brazil systematic country diagnostic, World Bank, Washington, DC.