Physica A 390 (2011) 2051–2066

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Identifying the community structure of the international-trade multi-network

Matteo Barigozzi [a], Giorgio Fagiolo [b,*], Giuseppe Mangioni [c]

[a] Department of Statistics, London School of Economics and Political Science, UK

[b] Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Pisa, Italy

[c] Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, University of Catania, Italy

ARTICLE INFO

Article history:
Received 22 September 2010
Received in revised form 12 January 2011
Available online 12 February 2011

Keywords:
Networks
Community structure
International-trade multi-network
Normalized mutual information

ABSTRACT

We study the community structure of the multi-network of commodity-specific trade relations among world countries over the 1992–2003 period. We compare structures across commodities and time by means of the normalized mutual information index (NMI). We also compare them with exogenous community structures induced by geography and regional trade agreements. We find that commodity-specific community structures are very heterogeneous and much more fragmented than that characterizing the aggregate ITN. This shows that the aggregate properties of the ITN may result (and be very different) from the aggregation of very diverse commodity-specific layers of the multi-network. We also show that commodity-specific community structures, especially those related to the chemical sector, are becoming more and more similar to the aggregate one. Finally, our findings suggest that geography-induced partitions of our set of countries are much more correlated with observed community structures than partitions induced by regional-trade agreements. This result strengthens previous findings from the empirical literature on trade.

1. Introduction

In the last years there was a surge of interest in the study of international-trade issues from a complex-network perspective [1–11]. Many contributions have explored the evolution over time of the topological properties of the aggregate International Trade Network (ITN), aka the World Trade Web (WTW), defined as the graph of total import/export relationships between world countries in a given year. More recently, a number of papers have instead begun to investigate the multi-network of trade [12,13], where a commodity-specific approach is followed to unfold the aggregate ITN in many layers, each one representing import and export relationships between countries for a given commodity class (cf. also Refs. [14,15] and the pioneering work of Paul Slater, cf. Refs. [16,17]).

In this paper, we explore further the topological architecture of the multi-network of international trade studying, for the first time, its community structure (see Ref. [18] for an overview). Detecting the community structure of the ITN and how it correlates with country-specific variables and geography (e.g., distances between countries) is crucial from an international-trade perspective. Indeed, finding communities in the ITN means identifying clusters of countries that carry tightly interrelated trade linkages among them, while being relatively less interconnected with countries outside the cluster. To date, only two papers have been trying to explore the community structure of the ITN [19,20]. However, they have only studied the aggregate ITN, i.e. the network obtained from total import/export relations between countries irrespective of the specific commodity traded. By focusing on the aggregate ITN only, one indeed neglects the fact that countries actually trade different lines of products and mostly employ imported goods either as inputs to the production process, or as consumption

* Corresponding author.
E-mail addresses: M.Barigozzi@lse.ac.uk (M. Barigozzi), giorgio.fagiolo@sssup.it (G. Fagiolo), Giuseppe.Mangioni@dieei.unict.it (G. Mangioni).

0378-4371/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.02.004

goods. Therefore, identifying clusters of countries from a multi-network perspective may be relevant to better understand what are the countries in the world that tend to trade the same group of products over time and, in turn, uncovering some stylized facts about the actual input–output and supply-demand interdependencies between countries. This may be relevant to predict, for example, to what extent a negative shock hitting a particular industry in a certain region of the world (or in a cluster) may spread and affect the same industry (or closely related ones) in another region of the world (or in another cluster).

Here, we begin addressing this issue by detecting the community structure characterizing the commodity-specific ITN over the period 1992–2003 ($T = 12$ years). We employ data about 162 countries and 97 commodities (2-digit disaggregation), to build a sequence of T multi ITNs. We begin by focusing on the 14 top-traded and economically relevant commodities, identifying the community structure of each layer (i.e. groups of countries that trade a given commodity). We then compare commodity-specific community structures with a number of properly-specified community benchmarks. These benchmarks are partitions of the set of our 162 countries obtained from: (i) the community structures of the aggregate ITN; (ii) exogenous continental or macro-areas classifications; (ii) the network of geographical closeness (i.e., the inverse of geographical distance between countries); (iii) the regional trade agreement (RTA) network. The main question we ask is whether (and how) commodity-specific community structures are similar to, or differ from, those detected in the benchmark networks. Comparisons are made using the normalized mutual information index (NMI), which is a measure of how close two partitions of the same set of N units are [21]. Understanding whether community structures detected at the commodity-specific level are similar to — or different from — those detected in the benchmark networks can shed further light on the topological architecture of the ITN. For example, comparing aggregate and commodity-specific community structures may tell us whether the community structure that we observe at the aggregate trade level can be explained by the aggregation of heterogeneous community structures or, conversely, trade community formation is not affected too much by the type of commodity traded. Similarly, comparing trade-induced communities with those obtained through geographically-induced networks may help us to understand the extent to which the formation of trade communities is related to geographical distance (as a proxy of trade resistance factors, e.g. trade fees).

The rest of the paper is organized as follows. Section 2 describes the databases that we employ in our exercises. Section 3 explains the community detection method that we use in this work. Section 4 discusses our main results. Concluding remarks are in Section 5.

2. Data and definitions

We employ bilateral trade flows data from the United Nations Commodity Trade Database (UN-COMTRADE; see http://comtrade.un.org/). We build a balanced panel of $N = 162$ countries (see Table 1) for which we have commodity-specific imports and exports flows from 1992 to 2003 ($T = 12$ years) in current US dollars. Trade flows are reported for $C = 97$ (2-digit) different commodities, classified according to the Harmonized System 1996 (HS1996; http://www.wcoomd.org/).[1]

We employ the database to build a time sequence of weighted, directed multi-graphs (multi-networks, henceforth) of trade where the N nodes are world countries and directed links represent the value of exports of a given commodity in each year $t = 1992, \dots, 2003$.[2] As a result, we have a time sequence of T multi-networks of international trade, each characterized by C layers (or links of C different colors). Each layer $c = 1, \dots, C$ represents exports between countries for commodity c and can be characterized by a $N \times N$ weight matrix X_t^c. Its generic entry $x_{ij,t}^c$ corresponds to the value of exports of commodity c from country i to country j in year t. We consider directed networks, therefore in general $x_{ij,t}^c \neq x_{ji,t}^c$. The aggregate weighted, directed ITN is obtained by simply summing up all commodity-specific layers. The entries of its weight matrices X_t read:

$$x_{ij,t} = \sum_{c=1}^{C} x_{ij,t}^c, \quad t = 1992, \dots, 2003. \tag{1}$$

For the sake of exposition, we shall focus on the most important commodity networks. Table 2 shows the ten most-traded commodities in 2003, ranked according to the total value of trade. Notice that they account, together, for 56% of total world trade and that the 10 most-traded commodities feature also the highest values of trade-value per link (i.e. ratio between total trade and total number of links in the commodity-specific network). In addition to these 10 trade-relevant commodities, we shall also focus on other 4 classes (cereals, cotton, coffee/tea and arms), which are less traded but more relevant in economics terms. The 14 commodities considered account together for 57% of world trade in 2003.[3]

1 The choice of a 2-digit breakdown of the data may be considered insufficient to clearly identify homogeneous product lines, but it has been made because in the HS classification system there is not a unique way to further disaggregate flows by commodities at a higher number of digits. Notice, however, that network analyses often face a trade-off between the need for a finer disaggregation and the very possibility to obtain connected graphs: typically, as soon as 3 or 4 digit data are considered, the resulting graphs easily become not connected, with the size of the largest connected component quickly decreasing.

2 Since, as always happens in trade data, exports from country i to country j are reported twice (according to the reporting country—importer or exporter) and sometimes the two figures do not match, we follow Ref. [22] and only employ import flows. For the sake of exposition, however, we follow the flow of goods and we treat imports from j to i as exports from i to j.

3 We refer the reader to Ref. [13] for a thorough analysis of the topological properties of this database from a multi-network perspective.

Table 1
Countries, continents and macro-areas.

ID	Country	Continent	Macro area
1	Albania	Europe	Eastern Europe
2	Algeria	Africa	North Africa and Middle East
3	Angola	Africa	Sub-Saharan Africa
4	Antigua Barbuda	America	Latin and Central America
5	Argentina	America	Latin and Central America
6	Armenia	Europe	Eastern Europe
7	Australia	Oceania	Oceania
8	Austria	Europe	Periphery Europe
9	Azerbaijan	Europe	Eastern Europe
10	Bahamas	America	Latin and Central America
11	Bahrain	Asia	North Africa and Middle East
12	Bangladesh	Asia	South and East Asia
13	Barbados	America	Latin and Central America
14	Belarus	Europe	Eastern Europe
15	Belgium–Luxembourg	Europe	Core Europe
16	Belize	America	Latin and Central America
17	Benin	Africa	Sub-Saharan Africa
18	Bhutan	Asia	South and East Asia
19	Bolivia	America	Latin and Central America
20	Brazil	America	Latin and Central America
21	Brunei	Asia	South and East Asia
22	Bulgaria	Europe	Eastern Europe
23	Burkina Faso	Africa	Sub-Saharan Africa
24	Burundi	Africa	Sub-Saharan Africa
25	Cambodia	Asia	South and East Asia
26	Cameroon	Africa	Sub-Saharan Africa
27	Canada	America	North America
28	Cape Verde	Africa	Sub-Saharan Africa
29	Central African Rep	Africa	Sub-Saharan Africa
30	Chad	Africa	Sub-Saharan Africa
31	Chile	America	Latin and Central America
32	China	Asia	South and East Asia
33	Colombia	America	Latin and Central America
34	Comoros	Africa	Sub-Saharan Africa
35	Congo	Africa	Sub-Saharan Africa
36	Costa Rica	America	Latin and Central America
37	Cote d'Ivoire	Africa	Sub-Saharan Africa
38	Croatia	Europe	Eastern Europe
39	Cyprus	Europe	Periphery Europe
40	Dem Rep Congo	Africa	Sub-Saharan Africa
41	Denmark	Europe	Periphery Europe
42	Djibouti	Africa	Sub-Saharan Africa
43	Dominica	America	Latin and Central America
44	Dominican Rep	America	Latin and Central America
45	Ecuador	America	Latin and Central America
46	Egypt	Africa	North Africa and Middle East
47	El Salvador	America	Latin and Central America
48	Equatorial Guinea	Africa	Sub-Saharan Africa
49	Fiji	Oceania	Oceania
50	Finland	Europe	Periphery Europe
51	Fmr Ethiopia	Africa	Sub-Saharan Africa
52	France	Europe	Core Europe
53	Gabon	Africa	Sub-Saharan Africa
54	Gambia	Africa	North Africa and Middle East
55	Germany	Europe	Core Europe
56	Ghana	Africa	Sub-Saharan Africa
57	Greece	Europe	Periphery Europe
58	Grenada	America	Latin and Central America
59	Guatemala	America	Latin and Central America
60	Guinea	Africa	Sub-Saharan Africa
61	Guinea Bissau	Africa	Sub-Saharan Africa
62	Guyana	America	Latin and Central America
63	Haiti	America	Latin and Central America
64	Honduras	America	Latin and Central America
65	Hong Kong	Asia	South and East Asia
66	Hungary	Europe	Eastern Europe
67	Iceland	Europe	Periphery Europe
68	India	Asia	South and East Asia
69	Indonesia	Asia	South and East Asia

(*continued on next page*)

Table 1 (*continued*)

ID	Country	Continent	Macro area
70	Iran	Asia	Central Asia
71	Ireland	Europe	Periphery Europe
72	Israel	Asia	North Africa and Middle East
73	Italy	Europe	Core Europe
74	Jamaica	America	Latin and Central America
75	Japan	Asia	South and East Asia
76	Jordan	Asia	North Africa and Middle East
77	Kazakstan	Asia	Central Asia
78	Kenya	Africa	Sub-Saharan Africa
79	Kiribati	Oceania	Oceania
80	Korea	Asia	South and East Asia
81	Kuwait	Asia	North Africa and Middle East
82	Kyrgyzstan	Asia	Central Asia
83	Laos	Asia	South and East Asia
84	Latvia	Europe	Eastern Europe
85	Lebanon	Asia	North Africa and Middle East
86	Libya	Africa	North Africa and Middle East
87	Lithuania	Europe	Eastern Europe
88	Madagascar	Africa	Sub-Saharan Africa
89	Malawi	Africa	Sub-Saharan Africa
90	Malaysia	Asia	South and East Asia
91	Maldives	Asia	South and East Asia
92	Mali	Africa	Sub-Saharan Africa
93	Mauritania	Africa	Sub-Saharan Africa
94	Mauritius	Africa	Sub-Saharan Africa
95	Mexico	America	North America
96	Moldova	Europe	Eastern Europe
97	Mongolia	Asia	Central Asia
98	Morocco	Africa	North Africa and Middle East
99	Mozambique	Africa	Sub-Saharan Africa
100	Myanmar	Asia	South and East Asia
101	Nepal	Asia	South and East Asia
102	Netherlands	Europe	Core Europe
103	New Zealand	Oceania	Oceania
104	Nicaragua	America	Latin and Central America
105	Niger	Africa	Sub-Saharan Africa
106	Nigeria	Africa	Sub-Saharan Africa
107	Norway	Europe	Periphery Europe
108	Oman	Asia	North Africa and Middle East
109	Pakistan	Asia	Central Asia
110	Panama	America	Latin and Central America
111	Papua New Guinea	Oceania	Oceania
112	Paraguay	America	Latin and Central America
113	Peru	America	Latin and Central America
114	Philippines	Asia	South and East Asia
115	Poland	Europe	Eastern Europe
116	Portugal	Europe	Periphery Europe
117	Qatar	Asia	North Africa and Middle East
118	Romania	Europe	Eastern Europe
119	Russia	Europe	Eastern Europe
120	Rwanda	Africa	Sub-Saharan Africa
121	Saint Kitts and Nevis	America	Latin and Central America
122	Saint Lucia	America	Latin and Central America
123	Saint Vincent and the Grenadines	America	Latin and Central America
124	Samoa	Oceania	Oceania
125	Sao Tome and Principe	Africa	Sub-Saharan Africa
126	Saudi Arabia	Asia	North Africa and Middle East
127	Senegal	Africa	Sub-Saharan Africa
128	Seychelles	Africa	Sub-Saharan Africa
129	SierraLeone	Africa	Sub-Saharan Africa
130	Singapore	Asia	South and East Asia
131	Slovenia	Europe	Eastern Europe
132	South African Customs Union	Africa	Sub-Saharan Africa
133	Solomon Isds	Oceania	Oceania
134	Spain	Europe	Core Europe
135	Sri Lanka	Asia	South and East Asia
136	Sudan	Africa	Sub-Saharan Africa
137	Suriname	America	Latin and Central America
138	Sweden	Europe	Periphery Europe
139	Switzerland	Europe	Core Europe
140	Syria	Asia	North Africa and Middle East

Table 1 (*continued*)

ID	Country	Continent	Macro area
141	Tajikistan	Asia	Central Asia
142	Tanzania	Africa	Sub-Saharan Africa
143	Thailand	Asia	South and East Asia
144	Togo	Africa	Sub-Saharan Africa
145	Tonga	Oceania	Oceania
146	Trinidad Tobago	America	Latin and Central America
147	Tunisia	Africa	North Africa and Middle East
148	Turkey	Asia	North Africa and Middle East
149	Turkmenistan	Asia	Central Asia
150	Uganda	Africa	Sub-Saharan Africa
151	Ukraine	Europe	Eastern Europe
152	United Arab Emirates	Asia	North Africa and Middle East
153	United Kingdom	Europe	Core Europe
154	Uruguay	America	Latin and Central America
155	USA	America	North America
156	Uzbekistan	Asia	Central Asia
157	Vanuatu	Oceania	Oceania
158	Venezuela	America	Latin and Central America
159	Viet Nam	Asia	South and East Asia
160	Yemen	Asia	North Africa and Middle East
161	Zambia	Africa	Sub-Saharan Africa
162	Zimbabwe	Africa	Sub-Saharan Africa

Table 2
The 14 most relevant commodity classes in year 2003 in terms of total-trade value (USD), trade value per link (USD), and share of world aggregate trade.

Code	Commodity	Value (USD)	Value per link (USD)	% of Aggregate trade (%)
83	Nuclear reactors, boilers, machinery and mechanical appliances; parts thereof	5.67×10^{11}	6.17×10^{7}	11.37
84	Electric machinery, equipment and parts; sound equipment; television equipment	5.58×10^{11}	6.37×10^{7}	11.18
27	Mineral fuels, mineral oils and products of their distillation; bitumin substances; mineral wax	4.45×10^{11}	9.91×10^{7}	8.92
86	Vehicles (not railway, tramway, rolling stock); parts and accessories	3.09×10^{11}	4.76×10^{7}	6.19
89	Optical, photographic, cinematographic, measuring, checking, precision, medical or surgical instruments/apparatus; parts and accessories	1.78×10^{11}	2.48×10^{7}	3.58
39	Plastics and articles thereof	1.71×10^{11}	2.33×10^{7}	3.44
29	Organic chemicals	1.67×10^{11}	3.29×10^{7}	3.35
30	Pharmaceutical products	1.4×10^{11}	2.59×10^{7}	2.81
72	Iron and steel	1.35×10^{11}	2.77×10^{7}	2.70
71	Pearls, precious stones, metals, coins, etc.	1.01×10^{11}	2.41×10^{7}	2.02
10	Cereals	3.63×10^{10}	1.28×10^{7}	0.73
52	Cotton, including yarn and woven fabric thereof	3.29×10^{10}	6.96×10^{6}	0.66
09	Coffee, tea, mate and spices	1.28×10^{10}	2.56×10^{6}	0.26
92	Arms and ammunition, parts and accessories thereof	4.31×10^{9}	2.46×10^{6}	0.09
ALL	Aggregate	4.99×10^{12}	3.54×10^{8}	100.00

We also employ data about regional trade agreements (RTAs) between world countries taken from the World Trade Organization (WTO) website.[4] We build a weighted undirected network with weight matrix $M_t = \{m_{ij,t}\}$ where nodes are countries and a link is weighted according to the number $m_{ij,t}$ of RTAs — free, multilateral and/or bilateral — in place between the two countries i and j at year t (cf. also Ref. [20]). This sequence of networks may be interpreted as an indicator of how intense are trade agreements between countries over time, i.e. how close countries are in the RTA space.[5] It is well-known from the empirical literature on trade that RTAs are an important determinant of trade flows [23].

Finally, we employ data on geographical positions of countries to build geographically-related networks and partitions (see Table 1). First, we consider a standard partition of our 162 countries in 5 continents (America, Africa, Europe, Asia, Australia). Second, we partition countries in a slightly more disaggregated classification using the following macro

4 See http://www.wto.org/.

5 RTAs may strongly differ in the value of trade induced by the agreement. However, we do not employ a RTA matrix weighted by that value (total trade between countries involved by the agreement) for two related reasons. First, it is not clear from an empirical perspective the net impact on total trade of a given RTA. Second, by weighting RTAs with trade values we would end up with a RTA network very much correlated with the ITN. This would bias our findings.

geographical areas: Core EU, Periphery EU, Eastern Europe, North and Central America, South America, South and East Asia, Central Asia, North Africa and Middle East, Sub-Saharan Africa, Oceania. These are obtained from the United Nations geographical subregions classification, cf. also Ref. [13]. Third, we build a geographically-related weighted undirected network with weights $s_{ij} = 1/f(d_{ij})$, where $f(\cdot)$ is a strictly increasing function and d_{ij} are the geographical distances between the most populated cities of country i and country j.[6] We have investigated several alternative functional forms for f, including $f(x) = x$, $f(x) = \sqrt{x}$, $f(x) = \log(x)$ and $f(x) = \exp(x)$. Since results are not dramatically affected by the choice of f, in what follows we will present our findings for $s_{ij} = 1/d_{ij}$. We employ the resulting matrix $S = \{s_{ij}\}$ as a weighted undirected network of geographical closeness between countries, i.e. as the network conveying information on how close countries are in the geographical space. Notice that, traditionally, geographical distance between countries is interpreted as a proxy of all factors that impose some resistance to free trade (transport costs, fees, etc.).

3. Community detection and comparison

It has been observed that many real networks exhibit a concentration of links within a special groups of nodes called communities (or clusters or modules). Such a structural property of a network has also been linked to the presence of sub-modules whose nodes have some functional property in common. Therefore, the detection of the community structure of a given network could help to discover some hidden feature of its topological architecture.

Despite the intuitive concept of community, a precise definition of what a community is represents a challenging issue (see Ref. [18]). In this paper we adopt the well-known formulation given in Ref. [24]: a subgraph is a community if the number of links (or, more generally, the intensity of interactions) among nodes in the subgraph is higher than what would be expected in an equivalent network with links (and intensities) placed at random. This definition implies the choice of a so-called "null model", i.e. a model of network to which the observed network can be statistically compared in order to assert the existence of any degree of modularity. The most used null model is a random network with the same number of nodes, the same number of links and the same degree distribution as in the original network, but with links among nodes randomly placed. Based on these concepts, a function called modularity that gives a measure of the quality of a given network partition into communities has been introduced in Ref. [24]. The modularity function has been further extended in Ref. [25] to the case of weighted directed networks as reported in the following:

$$Q = \frac{1}{W}\sum_{ij}\left[w_{ij} - \frac{w_i^{\text{out}} w_j^{\text{in}}}{W}\right]\delta_{c_i,c_j} \tag{2}$$

where w_{ij} is the weight of the link between i and j, $w_i^{\text{out}} = \sum_j w_{ij}$ and $w_j^{\text{in}} = \sum_i w_{ij}$ are respectively the output and input strengths of nodes i and j, $W = \sum_i \sum_j w_{ij}$ is the total strength of the network and δ_{c_i,c_j} is 1 if nodes i and j are in the same community and 0 otherwise.

In this paper communities are uncovered by optimizing the modularity function in Eq. (2). The optimization of Q is performed by using a tabu search algorithm.[7] We shall go back to some critical remarks on the use of modularity-based community-detection algorithms in the concluding section.

As discussed in Section 1, one of the contributions of this paper is to compare commodity-specific community structures with a proper number of community benchmarks (as detailed in the next section). To compare community partitions we use the *normalized mutual information* (NMI) measure, as introduced in Ref [21]. To define the NMI index, the confusion matrix plays a crucial role. Given two community partitions $\mathcal{P}_A$ and $\mathcal{P}_B$, the confusion matrix $\mathcal{N}$ is defined as a matrix whose N_{ij}-th element is the number of nodes in the community i of the partition $\mathcal{P}_A$ that appear in the community j of the partition $\mathcal{P}_B$. The NMI is defined as:

$$\text{NMI}(\mathcal{P}_A, \mathcal{P}_B) = \frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B} N_{ij}\log\left(\frac{N_{ij}N}{N_{i.}N_{.j}}\right)}{\sum_{i=1}^{C_A} N_{i.}\log\left(\frac{N_{i.}}{N}\right) + \sum_{j=1}^{C_B} N_{.j}\log\left(\frac{N_{.j}}{N}\right)} \tag{3}$$

where C_A and C_B are respectively the number of communities in $\mathcal{P}_A$ and $\mathcal{P}_B$, $N_{i.} = \sum_j N_{ij}$, $N_{.j} = \sum_i N_{ij}$ and $N = \sum_i \sum_j N_{ij}$. The NMI index is equal to 1 if $\mathcal{P}_A$ and $\mathcal{P}_B$ are identical and assumes a value of 0 if the two partitions are independent.

4. Results

4.1. Detecting the community structure of the multi ITN

We begin by studying the connectivity of the multi ITN and the size and concentration of its community structures. All results refer to the aggregate ITN and to the 14 commodity-specific layers as defined in Section 2. In Table 3 we show the

[6] Results are robust to alternative distance measures. Data are definitions are available at the URL: http://www.cepii.fr/.

[7] Tabu search [26] is a local-search optimization method that enhances the performance of local search by using memory: once a potential solution has been determined, it is marked as "taboo", so that the algorithm does not visit that possibility repeatedly.

Table 3
Density of aggregate ITN and relative density of commodity specific ITNs with respect to the aggregate.

Commodity year	All	Coffee tea (%)	Cereals (%)	Mineral fuels (%)	Organic chem. (%)	Pharma. prod. (%)	Plastics (%)	Cotton (%)	Precious stones (%)	Iron steel (%)	Nuclear reac. (%)	Electric mach. (%)	Vehicles (%)	Optical inst. (%)	Arms (%)
1992	0.2260	38	17	30	36	33	44	36	34	33	58	55	42	45	14
1993	0.2832	37	16	29	36	33	45	36	34	32	59	56	42	45	13
1994	0.3602	36	18	30	36	34	47	38	33	32	60	57	43	47	14
1995	0.4199	34	18	30	35	36	46	36	30	33	60	57	44	47	13
1996	0.4553	35	19	30	35	35	47	35	29	33	61	58	45	46	13
1997	0.4925	33	19	30	35	36	48	34	29	33	61	59	45	47	13
1998	0.5118	34	20	30	35	36	48	34	28	33	62	59	45	48	12
1999	0.5297	33	20	30	35	37	49	34	28	33	62	59	45	47	12
2000	0.5406	33	19	30	35	37	50	34	28	33	63	60	45	48	12
2001	0.5570	33	19	31	35	38	50	33	27	33	63	60	45	48	12
2002	0.5334	33	20	32	35	38	51	33	28	33	64	61	46	49	12
2003	0.5400	35	20	32	36	38	52	34	30	35	65	62	46	51	12

Table 4
Size of the largest connected components (LCC).

Commodity year	All	Coffee tea	Cereals	Mineral fuels	Organic chem.	Pharma. prod.	Plastics	Cotton	Precious stones	Iron steel	Nuclear reac.	Electric mach.	Vehicles	Optical inst.	Arms
1992	162	145	111	127	138	129	151	147	151	144	161	161	154	157	90
1993	162	150	129	143	143	140	158	153	160	149	162	162	159	162	105
1994	162	155	133	148	152	152	160	156	160	158	162	162	161	160	111
1995	162	158	144	156	157	155	161	156	160	162	162	162	161	161	120
1996	162	158	145	156	153	153	161	157	158	158	162	162	159	162	129
1997	162	162	150	155	151	154	159	156	161	160	162	162	161	162	130
1998	162	161	151	156	157	152	160	157	160	158	162	162	160	162	132
1999	162	160	153	160	156	157	161	159	161	160	162	162	162	162	129
2000	162	160	150	157	154	157	162	160	158	161	162	162	161	162	137
2001	162	161	152	160	160	160	161	157	159	160	162	162	162	162	139
2002	162	160	150	160	158	158	161	158	158	157	162	162	162	162	139
2003	162	161	150	158	157	158	162	161	161	159	162	162	162	162	134

evolution of the density of the aggregate ITN (computed as the ratio between the number of existing links to the number of all possible links, i.e. $N(N-1)$) and the relative density of commodity-specific networks (relative to the density of the aggregate ITN). We observe a monotonic increase in time of the aggregate ITN density, whereas the relative densities remained almost constant over time in each commodity-specific ITN. This implies an increase in the absolute value of commodity-specific densities. We also observe a relatively high heterogeneity of relative densities across commodity networks, which are always and significantly smaller than the aggregate one. This signals that results obtained using the aggregate ITN may be very different from those obtained looking at single commodity-specific networks (see also below).

While the density measures the concentration of trade links in a network, the size of the largest connected component (LCC) measures its overall level of connectivity. In Table 4, we report the size of the LCCs of the aggregate and commodity specific ITNs. While the former is always a completely connected network, this is not always the case for disaggregated cases, see for example arms and cereals. Other commodities, including electronics, optics, plastics and coffee, show instead a large connectivity close to that of the aggregate ITN. This means that their contribution to overall connectivity is very strong. Notice also that for all the commodities we observe an increase in time in the size of the LCC, which is clearly a sign of the increase in the degree of integration of world trade. The largest changes in the size of LCC are observed for arms ($c = 92$), cereals ($c = 10$) and pharmaceutical products ($c = 30$).

We now detect the community structure of both aggregate and commodity-specific ITNs by maximizing a weighted-directed version of the modularity function (see Eq. (2)). We employ the community structure of the 12 yearly aggregate trade networks with weight matrix as in Eq. (1) as a first benchmark, in order to compare commodity-specific clusters with that obtained from the aggregate trade flows.

The number of communities that we identify in each year and network is shown in Table 5. To begin, notice how the aggregate ITN typically displays a smaller number of communities than most of commodity-specific networks, meaning that the latter are more fragmented as far as trade clusters are concerned. In addition, we also observe that the number of communities in the aggregate ITN steadily increases over time, whereas this is not the case for most commodity-specific trade networks. In general, it appears that the smaller the size of the LCC, the higher the number of communities one finds. However, if for every ITN we look at the correlation across time between size of LCC and number of communities, results are different. While for some commodities a larger LCC size implies fewer communities, for others the opposite holds. In the first group we find coffee and tea ($c = 9$), pharmaceutical products ($c = 30$), precious stones ($c = 71$), and electric machinery ($c = 84$). In the second group we find all other commodities and the aggregate ITN. This evidence points to the existence of a large degree of heterogeneity in the number of community structures across commodity-specific

Table 5
Number of communities.

Commodity year	All	Coffee tea	Cereals	Mineral fuels	Organic chem.	Pharma. prod.	Plastics	Cotton	Precious stones	Iron steel	Nuclear reac.	Electric mach.	Vehicles	Optical inst.	Arms
1992	2	6	5	4	3	6	3	5	5	3	3	4	3	2	7
1993	3	10	5	4	3	8	3	6	5	6	4	3	4	2	5
1994	3	8	8	5	3	6	4	4	5	5	3	2	4	2	6
1995	3	7	7	8	3	8	7	5	5	5	5	4	5	4	8
1996	3	5	7	6	4	6	5	6	5	6	4	3	5	3	9
1997	4	6	6	6	6	6	4	5	5	4	4	2	6	2	6
1998	4	6	6	6	4	6	6	5	4	4	5	5	4	2	7
1999	4	6	8	8	3	5	5	5	4	4	5	4	5	3	8
2000	4	7	6	5	4	6	6	6	4	4	5	4	6	3	6
2001	4	6	5	5	3	7	5	6	4	5	5	3	4	3	6
2002	4	6	6	4	4	6	7	5	4	5	2	3	5	3	8
2003	4	7	6	6	4	6	6	6	5	4	5	2	5	3	9

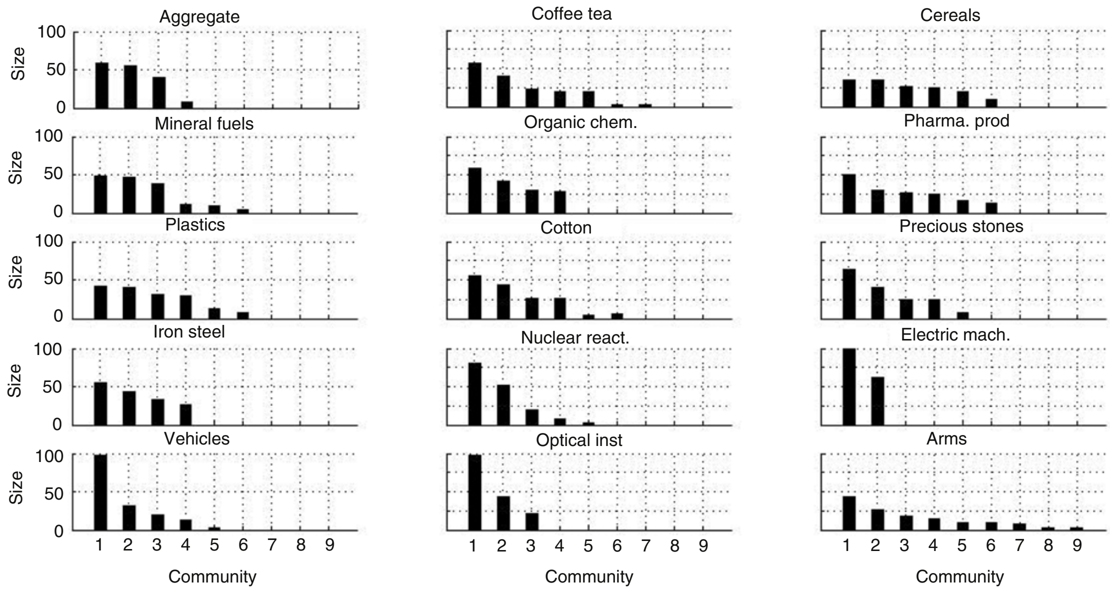


Fig. 1. Cluster-size distributions in 2003.

networks, suggesting that the results obtained in the case of the aggregate ITN hide a lot of variability in the community structure of commodity-specific networks. This result is in line with similar one obtained in Ref. [13], where it is shown that many properties of the aggregate ITN (e.g., log-normal distributions of link weights and node-specific characteristics like strength and clustering) are the sheer result of aggregating their counterparts across heterogeneous commodity-specific networks.

We now turn to a more detailed analysis of the community structure at a commodity-specific level. Fig. 1 shows the distributions of the cluster size in year 2003. Again, the shape of the distributions and their ranges vary a lot across commodities. The commodities that generate the most concentrated community structures are electric machinery ($c = 84$), optical instruments ($c = 89$), and vehicles ($c = 86$), i.e. products that require more scientific knowledge. To draw a more quantitative implication linking products and concentration of cluster-size distributions, we compute the normalized Herfindahl index (H), a synthetic measure of concentration of cluster size distributions. The index H, for a given commodity c and in a given year t, is defined as:

$$\mathrm{H}_t^c = \frac{1}{1-\frac{1}{N}} \left\{ \left[\sum_{i=1}^{n_{X_t^c,t}} \left(\frac{m_t^c(i)}{N} \right)^2 \right] - \frac{1}{N} \right\}, \quad t = 1992, \ldots, 2003, \tag{4}$$

where $n_{X_t^c,t}$ is the number of communities identified in the network X_t^c and $m_t^c(i)$ is the number of countries in the i-th community in year t for commodity c. The index ranges between 0 (no concentration at all) and 1 (maximum concentration). Table 6 reports the values of H_t^c for all networks and time periods. It is easy to notice that for the aggregate ITN there has been a decrease in concentration over time. This may be interpreted as a sign of the globalization process, as this pattern suggests that an increasing number of countries are participating in world trade over time. Indeed, while in 1992 we observe only 2 communities of about 80 countries each (one with Europe, Russia and Africa, the other with America and Asia), in 2003 a new community emerges, driven by China and India. At the commodity-specific level, an increase in H is observed

Table 6
Normalized Herfindal index H_t^c as a measure of concentration of communities.

Commodity year	All	Coffee tea	Cereals	Mineral fuels	Organic chem.	Pharma. prod.	Plastics	Cotton	Precious stones	Iron steel	Nuclear reac.	Electric mach.	Vehicles	Optical inst.	Arms
1992	0.50	0.16	0.12	0.16	0.25	0.14	0.30	0.30	0.30	0.27	0.38	0.41	0.49	0.48	0.08
1993	0.34	0.11	0.16	0.21	0.26	0.13	0.34	0.34	0.33	0.24	0.40	0.50	0.35	0.55	0.12
1994	0.46	0.15	0.14	0.26	0.36	0.18	0.33	0.33	0.26	0.30	0.44	0.51	0.35	0.52	0.16
1995	0.44	0.17	0.20	0.22	0.46	0.15	0.24	0.24	0.28	0.25	0.34	0.45	0.41	0.47	0.11
1996	0.41	0.24	0.16	0.20	0.38	0.18	0.37	0.37	0.27	0.28	0.35	0.40	0.44	0.45	0.14
1997	0.29	0.20	0.17	0.24	0.31	0.21	0.29	0.29	0.35	0.25	0.36	0.53	0.43	0.56	0.20
1998	0.29	0.19	0.19	0.21	0.30	0.18	0.25	0.25	0.37	0.25	0.24	0.48	0.45	0.54	0.16
1999	0.30	0.20	0.18	0.23	0.32	0.25	0.27	0.27	0.30	0.25	0.37	0.40	0.39	0.44	0.13
2000	0.28	0.22	0.20	0.22	0.25	0.24	0.27	0.27	0.28	0.26	0.34	0.44	0.36	0.39	0.20
2001	0.28	0.23	0.22	0.27	0.34	0.18	0.26	0.26	0.31	0.22	0.35	0.45	0.36	0.46	0.23
2002	0.29	0.24	0.21	0.27	0.30	0.25	0.24	0.24	0.28	0.21	0.53	0.41	0.46	0.45	0.15
2003	0.31	0.23	0.16	0.24	0.26	0.19	0.20	0.20	0.27	0.26	0.37	0.53	0.41	0.45	0.13

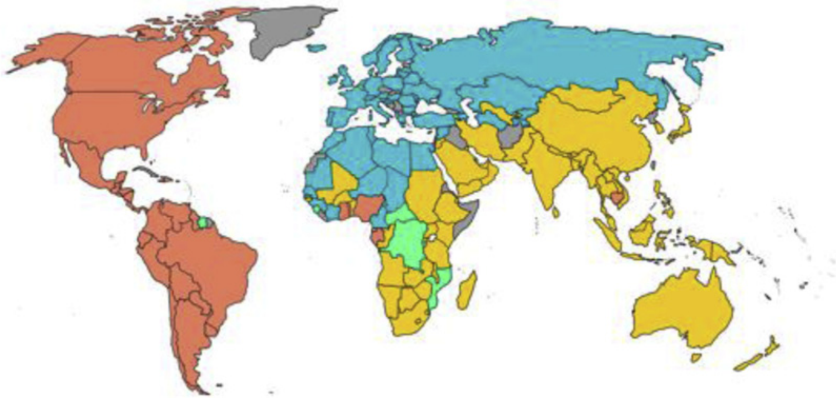


Fig. 2. World map showing communities of aggregate ITN in 2003. In gray countries not belonging to any community or for which no data are available.

also for coffee and tea ($c = 9$), mineral fuels ($c = 27$), pharmaceutical products ($c = 30$), arms ($c = 92$). However, for some other commodities we observe a decrease in H over time, see e.g. organic chemicals ($c = 29$), plastics ($c = 39$), and cotton ($c = 52$). This means that trade for those commodities has become less and less centralized and increasingly occurs among smaller and more dispersed groups of countries.

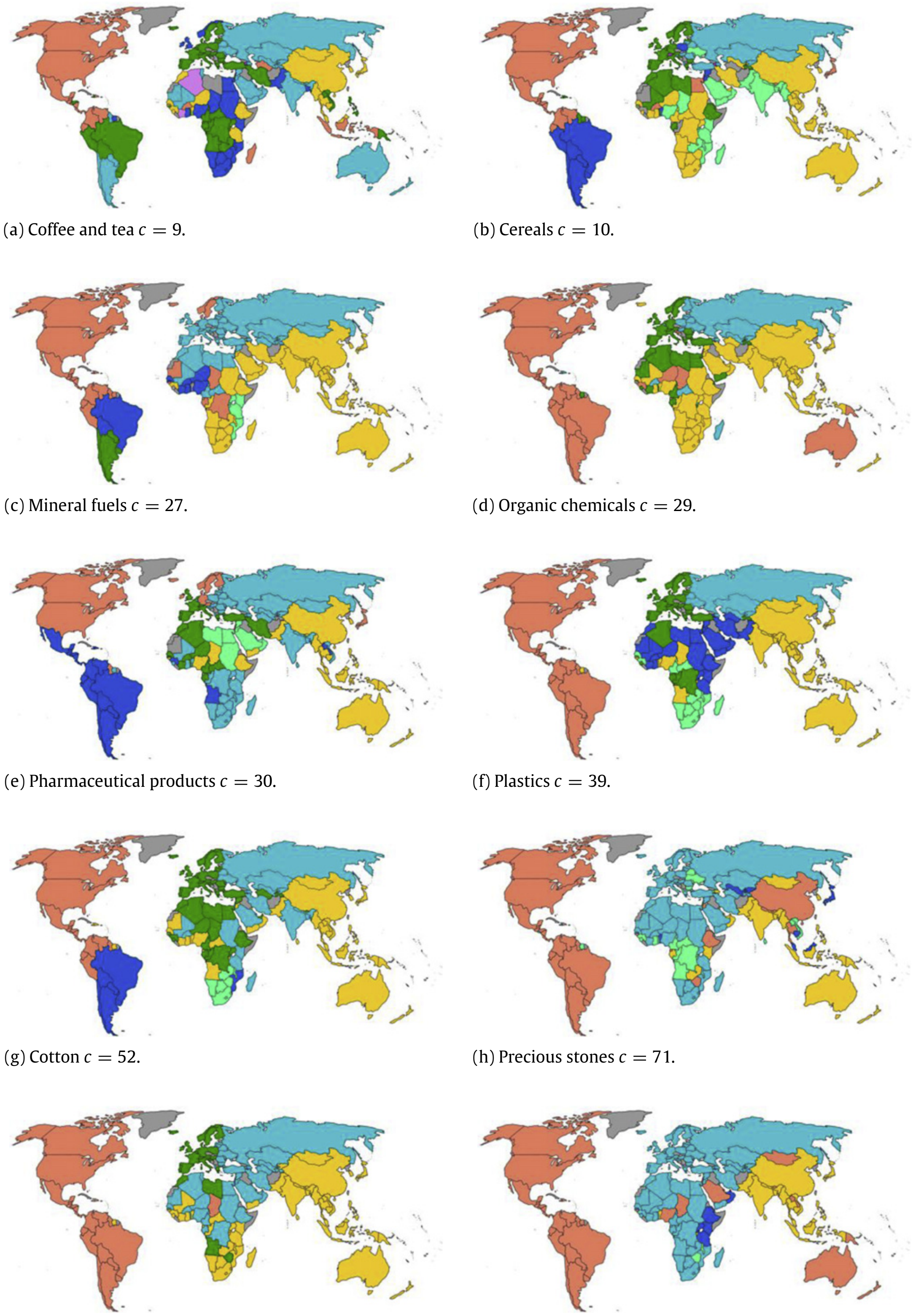
4.2. Describing trade communities

A useful way to visually describe community structure in the ITN is to employ colored world maps, where countries belonging to the same communities are associated to the same color. Figs. 2–4 report world maps depicting the community structure detected in 2003, for both the aggregate ITN and for the 14 commodity-specific networks.

Notice that this visual device also allows us to informally correlate community structures with geographical considerations (we shall go back to a more formal analysis of this issue below). For example, most of the networks studied exhibit the presence of an American cluster composed of US and Canada (and often linked to Latin America), a European cluster (sometimes connected to North Africa), an Asian cluster consisting of China (and in many cases of India, Indochina and Australia) and finally a Russian community (sometimes linked to the European cluster). Africa and Middle East are often split, independently of the commodity examined, among the other groups. This already suggests that geographical (and socio-political) factors are very important to explain the formation of community structure in the ITN.

Apart from the regularities above, commodity-specific community structures often differ in a relevant way among each other. In what follows, we highlight some economically-relevant features of aggregate and commodity-specific community structures in 2003. We focus on 7 commodity classes, those exhibiting the most economically relevant patterns (the remaining 7 classes did not show such explicit regularities). Due to the relatively strong persistence over time of ITNs topological architecture (see Refs. [11,13] for a discussion), similar considerations also hold for other years.

1. *Aggregate ITNs*: The world is divided in three major communities which follow a geographical pattern: (i) North and Latin America, (ii) Europe, Russia, and North Africa, (iii) China, India, Japan, Middle East, Australia and Sub-Saharan Africa. Two exceptions concern Africa: Nigeria and Ghana belong to the American community and we observe a minor separated community containing Belgium and the Democratic Republic of Congo, a former Belgian colony.
2. *Coffee and tea*: We identify two communities containing coffee drinking and producing countries: (i) Europe, Brazil, Peru, and Central African countries, (ii) North America, Central America, Colombia and Venezuela. We also identify two



(a) Coffee and tea $c = 9$.

(b) Cereals $c = 10$.

(c) Mineral fuels $c = 27$.

(d) Organic chemicals $c = 29$.

(e) Pharmaceutical products $c = 30$.

(f) Plastics $c = 39$.

(g) Cotton $c = 52$.

(h) Precious stones $c = 71$.

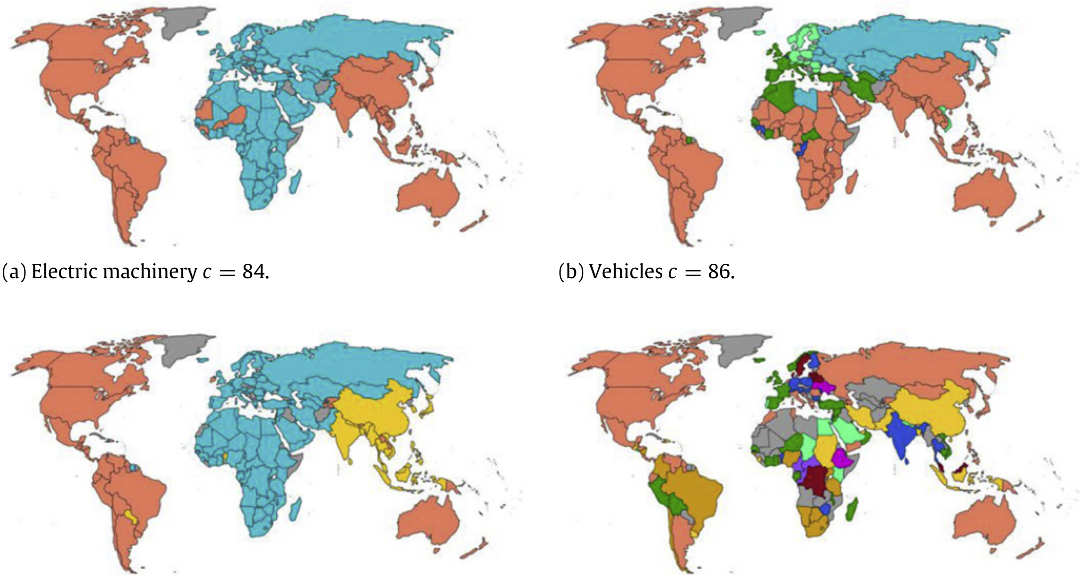
(i) Iron and steel $c = 72$.

(j) Nuclear reactors $c = 83$.

Fig. 3. World maps showing trade communities of commodity specific ITNs in 2003. In gray countries not belonging to any community or for which no data are available.

communities of mainly tea drinking and producing countries: (i) United Kingdom, South African and North-East African countries, Pakistan and Bangladesh. Finally, there exist two mixed communities, but probably more connected with the tea trade: (i) India, Middle East, Russia, Australia, Argentina, Chile, (ii) China, Japan, and Indochina.

3. *Cereals*: The big producers of cereals belong each to a separate community: (i) North America, (ii) South America, (iii) Russia. China and India are in separate communities, too. Finally, it is interesting to notice that Europe belongs to



(a) Electric machinery $c = 84$.

(b) Vehicles $c = 86$.

(c) Optical instruments $c = 89$.

(d) Arms $c = 92$.

Fig. 4. World maps showing trade communities of commodity specific ITNs in 2003. In gray countries not belonging to any community or for which no data are available.

yet another separate cluster. Despite being not a big producer, but a big consumer, Europe is not an open market for agricultural products. This finding may be linked to the protectionist agricultural policies of the European Community.

4. *Mineral fuels*: China and India have tight links with the Middle East, Europe has links with Russia and North Africa, Brazil with Nigeria, and North America with Norway, which is one of the largest oil producers in the world.
5. *Precious stones*: In this case America and China belong to the same community. Europe, Russia, North and South Africa are the members of the largest community. Interestingly, countries rich in diamonds such as the Democratic Republic of Congo, Angola, and Sierra Leone, belong to a unique community containing also Israel. Finally, Australia, Indonesia, and India belong to another cluster.
6. *Electric machinery*: There are only two communities strictly related to geographical distance. Indeed, countries within a community share common borders. One contains America, China, Japan, India and Australia. The other one contains Europe, Russia, Africa and the Middle East.
7. *Vehicles*: The world market for vehicles has a huge community containing America, China, Japan, India, Australia and almost all African countries. This may reflect the high diffusion of Japanese cars in Africa. Russia is still a closed market containing all former Soviet republics. Finally, Europe is divided in two communities, which have almost no members outside the continent: a finding that seems to reveal a protectionist market for vehicles in Europe.
8. *Arms*: The community structure for arms is highly fragmented and therefore difficult to interpret. Moreover, many countries seem not to belong to any community. Interestingly, these countries are often those where civil wars or in general social instability are most likely to be (or to have been) present. This is the case of Mozambique, Zambia, Angola, Guinea, Myanmar and Central Asian countries. It is unlikely that these countries do not participate in arms trade, but it is not surprising that our data do not reveal this, as probably that kind of trade relationships are not official. Finally, Africa is the most fragmented continent and almost all communities found contain some African countries.

4.3. Comparing community structures

In this section, we explore more quantitatively commodity-specific community structures by using the NMI index introduced in Section 3.

To begin with, we ask to what extent community structures are stable over the time interval considered. To do that, we compare the partitions obtained at time t and $t + 1$ for $t = 1992, \ldots, 2002$. More precisely, for the aggregate ITN and for any c, we compute the quantity $\mathrm{NMI}(\mathcal{P}_t^c, \mathcal{P}_{t+1}^c)$, where $\mathcal{P}_t^c$ is the partition of our N countries in year t for commodity c. This gives a measure of stability over time of community structures (see Table 7). Notice how the smallest values of NMI (large community structure changes) are observed in the early 1990s. In more recent years, on the contrary, NMIs have been larger, meaning weaker changes in the composition of communities from year to year. If one instead compares partitions in 1992 with those in 2003 (i.e., one computes the quantity $\mathrm{NMI}(\mathcal{P}_{1992}^c, \mathcal{P}_{2003}^c)$), it turns out that the stronger changes are associated to coffee and tea ($c = 9$), pharmaceutical products ($c = 30$), and arms ($c = 92$). The most stable community structures are instead those of aggregate trade, plastics ($c = 39$), optical instruments ($c = 89$), mineral fuels ($c = 27$), iron and steel ($c = 72$), and cotton ($c = 52$). Notice also that, on average, the majority of commodity-specific community structures were less stable than that of the aggregate network. Again, this suggests a strong mismatch between aggregate and disaggregated properties.

Table 7
NMI when comparing the community structures along the time dimension.

Commodity year	All	Coffee tea	Cereals	Mineral fuels	Organic chem.	Pharma. prod.	Plastics	Cotton	Precious stones	Iron steel	Nuclear reac.	Electric mach.	Vehicles	Optical inst.	Arms
1992–2003	0.27	0.17	0.22	0.29	0.24	0.17	0.35	0.28	0.20	0.29	0.23	0.22	0.20	0.30	0.16
1992–1993	0.54	0.38	0.25	0.35	0.27	0.28	0.42	0.29	0.35	0.39	0.38	0.20	0.25	0.25	0.33
1993–1994	0.41	0.46	0.32	0.48	0.24	0.35	0.48	0.34	0.32	0.41	0.38	0.48	0.42	0.37	0.35
1994–1995	0.55	0.42	0.43	0.50	0.28	0.45	0.45	0.45	0.17	0.45	0.37	0.51	0.29	0.42	0.29
1995–1996	0.51	0.33	0.47	0.52	0.37	0.44	0.52	0.34	0.27	0.46	0.40	0.56	0.51	0.56	0.33
1996–1997	0.58	0.27	0.34	0.50	0.40	0.56	0.39	0.35	0.43	0.46	0.40	0.55	0.52	0.50	0.32
1997–1998	0.75	0.51	0.43	0.61	0.32	0.51	0.56	0.44	0.48	0.56	0.54	0.59	0.58	0.71	0.33
1998–1999	0.77	0.56	0.52	0.71	0.56	0.48	0.57	0.35	0.42	0.57	0.39	0.47	0.51	0.39	0.34
1999–2000	0.73	0.49	0.49	0.55	0.40	0.44	0.59	0.42	0.49	0.44	0.50	0.50	0.55	0.42	0.33
2000–2001	0.79	0.63	0.47	0.60	0.38	0.54	0.46	0.55	0.47	0.52	0.56	0.50	0.42	0.40	0.29
2001–2002	0.68	0.55	0.55	0.59	0.34	0.64	0.45	0.41	0.34	0.57	0.43	0.44	0.41	0.63	0.30
2002–2003	0.65	0.49	0.53	0.55	0.42	0.46	0.57	0.52	0.32	0.52	0.37	0.58	0.47	0.58	0.28

Table 8
NMI when comparing the community structures induced by aggregate ITN with commodity-specific ITNs.

Commodity year	Coffee tea	Cereals	Mineral fuels	Organic chem.	Pharma. prod.	Plastics	Cotton	Precious stones	Iron steel	Nuclear reac.	Electric mach.	Vehicles	Optical inst.	Arms
1992	0.06	0.11	0.18	0.12	0.09	0.16	0.21	0.10	0.14	0.23	0.21	0.25	0.21	0.05
1993	0.17	0.14	0.37	0.29	0.23	0.44	0.24	0.18	0.28	0.25	0.19	0.16	0.21	0.11
1994	0.11	0.15	0.23	0.31	0.18	0.32	0.28	0.13	0.18	0.35	0.35	0.22	0.40	0.12
1995	0.14	0.20	0.25	0.30	0.22	0.40	0.28	0.21	0.24	0.36	0.40	0.19	0.26	0.15
1996	0.05	0.17	0.39	0.32	0.22	0.46	0.24	0.09	0.25	0.39	0.41	0.40	0.28	0.21
1997	0.12	0.32	0.46	0.33	0.28	0.30	0.29	0.20	0.37	0.46	0.25	0.30	0.24	0.18
1998	0.18	0.25	0.54	0.33	0.34	0.39	0.27	0.18	0.42	0.42	0.29	0.37	0.30	0.16
1999	0.15	0.33	0.56	0.31	0.29	0.51	0.38	0.26	0.37	0.39	0.38	0.31	0.31	0.12
2000	0.23	0.25	0.43	0.34	0.32	0.34	0.41	0.25	0.29	0.46	0.41	0.35	0.28	0.17
2001	0.20	0.26	0.54	0.38	0.31	0.50	0.33	0.28	0.43	0.42	0.40	0.47	0.39	0.15
2002	0.24	0.33	0.54	0.29	0.28	0.35	0.39	0.21	0.35	0.28	0.33	0.28	0.28	0.16
2003	0.14	0.27	0.49	0.32	0.29	0.45	0.38	0.28	0.38	0.37	0.19	0.25	0.26	0.12

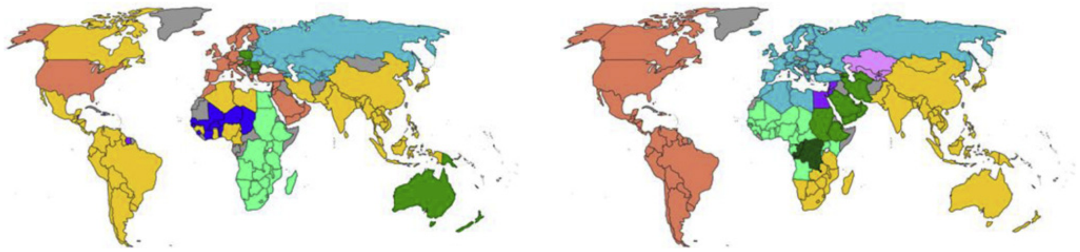
We then compare, in each given year, partitions associated to the aggregate ITN with those associated to commodity-specific networks by computing the quantities $\mathrm{NMI}(\mathcal{P}_t^{\mathrm{all}}, \mathcal{P}_t^c)$, where $\mathcal{P}_t^{\mathrm{all}}$ is the partition obtained from the aggregate ITN. This exercise is meant to ask more quantitatively the question whether the aggregate community structure can predict those obtained at the commodity-specific level well, or, to put it differently, the extent to which the community structure of any commodity-specific network contributes to shape (or is able to predict) that observed at the aggregate-trade level. Inspection of Table 8 shows that NMI values are increasing in time for almost all commodities. This means that commodity-specific community structures are becoming more and more similar to the aggregate one, i.e. that the role of all commodities in shaping the aggregated community structure has increased in time. In particular, we observe the largest increase in NMI from 1992 to 2003 for mineral fuels ($c = 27$), plastics ($c = 39$), iron and steel ($c = 72$), pharmaceutical products ($c = 30$), and organic chemicals ($c = 29$). Moreover, for all the years considered, mineral fuels and plastics appear to be the commodities whose community structure is the most similar to the aggregated one. Overall, a major role for the chemical sector emerges from these results, in that they are the ones whose country partition better mimics the aggregate one.

4.4. Community structure, geography, and trade agreements

We finally turn to study the extent to which community structures identified using trade data correlate with other economics-relevant data. To do that, we employ continental and macro-area geographical partitions, the geographical closeness matrix S, as well as the time-dependent RTA networks M_t. The entries of the symmetric and time-independent matrix S, to repeat, express a measure of geographical closeness between pairs of countries, computed as the inverse of geographical distance between their most populated cities. The entries of the symmetric but time-dependent matrices M_t contain, in a given year, the number of trade agreements currently in place between any two countries, irrespective of the type of RTA signed (bilateral, multilateral, commodity-specific, etc.). The underlying assumption is that the higher this number, the closer the two countries are in the RTA space (and thus, according to empirical findings, the larger their expected trade flows).

We apply to both S and M_t the community-detection algorithms explained in Section 3 and previously applied to trade matrices.[8] Therefore, we end up with three geographically-induced community partitions ($\mathcal{P}^{\mathrm{CONT}}$, $\mathcal{P}^{\mathrm{MACRO}}$, $\mathcal{P}^{\mathrm{GEO}}$) and 12

[8] Note that, strictly speaking, the null model featured in the definition of the modularity function is not the most appropriate one when applied to transformations of the distance matrix, as it does not preserve, e.g., triangular inequalities that the Euclidean metric satisfies. As we will see below, however, this does not seem to generate dramatic problems, as our main results also hold when we use exogenous geographical partitions instead of distance-based ones.



(a) RTAs. (b) Distances.

Fig. 5. World maps showing RTAs in 2003 and geographic communities.

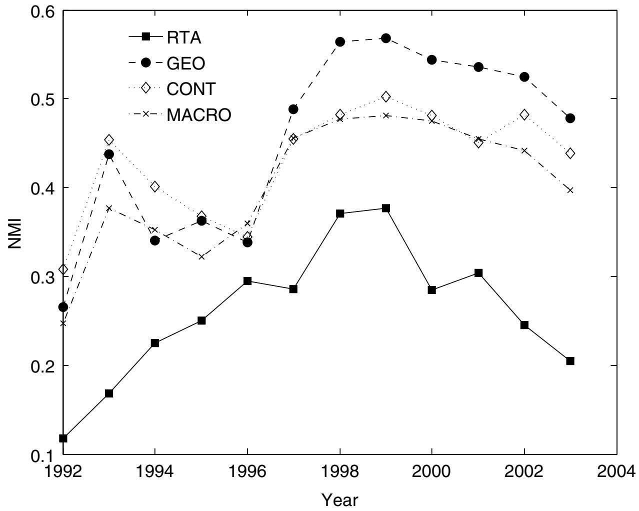


Fig. 6. NMI when comparing the community structures induced by the exogenous networks build using geographical distances (GEO), the partition of countries according to continents (CONT) or macro-areas (MACRO), and regional trade agreements data (RTA) with the community structures of aggregate trade.

time-dependent RTA-induced partitions ($\mathcal{P}_t^{\mathrm{RTA}}$). $\mathcal{P}^{\mathrm{CONT}}$ and $\mathcal{P}^{\mathrm{MACRO}}$ are the exogenous partitions of our 162 countries simply obtained by assigning each country to a continent or macro-area. $\mathcal{P}^{\mathrm{GEO}}$ and $\mathcal{P}_t^{\mathrm{RTA}}$ are the community structures that we get by maximizing modularity associated to the weighted undirected network defined by S and M_t, respectively. Their resulting partitions are visualized in the maps of Fig. 5. We find that $\mathcal{P}^{\mathrm{CONT}}$ and $\mathcal{P}^{\mathrm{MACRO}}$ turn out to be very similar to $\mathcal{P}^{\mathrm{GEO}}$ (with NMI values close to 0.7). This means that one can safely focus on $\mathcal{P}^{\mathrm{GEO}}$ as our relevant geography-induced partition. Notice that clusters in $\mathcal{P}^{\mathrm{GEO}}$ represent groups of countries that are geographically close, without using exogenously-determined partitions of countries (based on continents or macro-areas) and fully exploits the information coming from the distance matrix. Instead, the community structures in $\mathcal{P}_t^{\mathrm{RTA}}$ pick up clusters of countries that not only belong to free-trade or multilateral agreements (e.g. NAFTA, Mercosur, EU, etc.), but also signed additional bilateral agreements.

We first compare, using the NMI, the aggregate ITN community structure with those detected using geography or RTAs (see Fig. 6). We observe increasing NMIs across time until 2001 and a slight decrease afterwards. We also find more similarity between aggregate trade and geography-based communities with respect to communities determined by RTAs. Note that this result holds, as expected, irrespective of whether we employ continents, macro-areas, or distance-induced partitions obtained via modularity maximization. Thus, geographically-related factors seem to explain the patterns of global trade more than political determinants. Also, this result is more evident in the recent years after 2001. A possible explanation might be the global political crisis after 11th September 2001 that implied a slight decrease in global trade as a consequence of the wars in Iraq and Afghanistan.

When comparing community structures of commodity-specific ITNs with the partitions obtained from distance-induced geography and RTA data (see Tables 9 and 10), we find results similar to the aggregate case. In general, it is geography and not trade agreements that seems to correlate more with the observed patterns. Plastics ($c = 39$) and mineral fuels ($c = 27$) display the highest similarity with RTA communities. The same result holds when confronting trade communities with geographical data, but in addition we notice high NMIs also for iron and steel ($c = 72$) and cotton ($c = 52$). Again, this finding holds irrespective of whether we employ distance-based partitions or continental/macro-area exogenous breakdowns.

These results reinforce the traditional view put forth by standard gravity-equation trade empirics [23], which stresses the importance of geographical distance (as a proxy for trade resistance factors) in determining bilateral trade flows. Here, we show that geographical distance is important to predict not only the expected flow of a bilateral trade relationship (e.g., exports from country A to country B), but also the formation of trade communities, that is complicated trade structures

Table 9
NMI when comparing the community structures induced by geographical distances with commodity-specific ITNs.

Commodity year	Coffee tea	Cereals	Mineral fuels	Organic chem.	Pharma. prod.	Plastics	Cotton	Precious stones	Iron steel	Nuclear reac.	Electric mach.	Vehicles	Optical inst.	Arms
1992	0.24	0.26	0.41	0.31	0.29	0.37	0.30	0.21	0.30	0.25	0.24	0.29	0.26	0.21
1993	0.24	0.28	0.38	0.37	0.38	0.43	0.32	0.25	0.42	0.31	0.24	0.28	0.21	0.22
1994	0.23	0.38	0.47	0.27	0.34	0.43	0.30	0.21	0.36	0.29	0.28	0.38	0.27	0.23
1995	0.26	0.41	0.52	0.30	0.33	0.54	0.27	0.31	0.43	0.34	0.31	0.31	0.31	0.30
1996	0.18	0.32	0.43	0.26	0.31	0.38	0.32	0.22	0.43	0.30	0.34	0.32	0.33	0.28
1997	0.23	0.42	0.48	0.34	0.31	0.42	0.40	0.27	0.52	0.43	0.31	0.31	0.27	0.26
1998	0.21	0.37	0.52	0.39	0.39	0.43	0.35	0.22	0.41	0.41	0.31	0.35	0.31	0.23
1999	0.20	0.43	0.56	0.30	0.29	0.47	0.48	0.32	0.44	0.42	0.44	0.26	0.36	0.21
2000	0.30	0.36	0.45	0.40	0.40	0.44	0.48	0.33	0.37	0.43	0.43	0.33	0.30	0.21
2001	0.26	0.39	0.52	0.33	0.36	0.53	0.44	0.35	0.46	0.38	0.45	0.38	0.41	0.17
2002	0.27	0.41	0.53	0.31	0.32	0.47	0.44	0.25	0.47	0.30	0.35	0.33	0.32	0.21
2003	0.20	0.39	0.50	0.33	0.37	0.57	0.44	0.29	0.52	0.36	0.27	0.33	0.32	0.19

Table 10
NMI when comparing the community structures induced by regional trade agreements with commodity-specific ITNs.

Commodity year	Coffee tea	Cereals	Mineral fuels	Organic chem.	Pharma. prod.	Plastics	Cotton	Precious stones	Iron steel	Nuclear reac.	Electric mach.	Vehicles	Optical inst.	Arms
1992	0.15	0.17	0.20	0.13	0.20	0.17	0.22	0.12	0.19	0.16	0.15	0.14	0.13	0.15
1993	0.19	0.19	0.21	0.13	0.22	0.15	0.24	0.13	0.23	0.19	0.14	0.15	0.12	0.15
1994	0.16	0.27	0.24	0.21	0.26	0.21	0.19	0.14	0.26	0.16	0.20	0.19	0.16	0.17
1995	0.18	0.29	0.37	0.18	0.32	0.38	0.23	0.18	0.31	0.30	0.27	0.25	0.24	0.20
1996	0.11	0.28	0.33	0.24	0.27	0.31	0.26	0.16	0.26	0.25	0.30	0.29	0.21	0.23
1997	0.13	0.33	0.30	0.27	0.24	0.26	0.24	0.15	0.22	0.28	0.17	0.23	0.17	0.14
1998	0.27	0.31	0.36	0.22	0.36	0.42	0.28	0.14	0.27	0.32	0.25	0.30	0.20	0.21
1999	0.19	0.33	0.37	0.20	0.32	0.41	0.30	0.23	0.31	0.24	0.24	0.25	0.22	0.21
2000	0.25	0.29	0.35	0.22	0.25	0.33	0.33	0.22	0.26	0.26	0.30	0.24	0.15	0.17
2001	0.20	0.25	0.31	0.18	0.27	0.34	0.27	0.21	0.28	0.29	0.26	0.25	0.22	0.17
2002	0.20	0.30	0.30	0.19	0.25	0.34	0.31	0.20	0.29	0.17	0.23	0.25	0.22	0.19
2003	0.16	0.24	0.32	0.20	0.29	0.33	0.24	0.16	0.24	0.25	0.21	0.26	0.19	0.18

multilaterally involving groups of countries. On the other hand, our findings contribute to the discussion related to the impact of international agreements on world trade and seem to go in the direction of Ref. [27], which shows that there is no evidence that the WTO has increased international trade.

5. Concluding remarks

In this paper, we have provided a first exploratory study of the community structure of commodity-specific trade networks from 1992–2003. After recovering the optimal partition of countries, we compare commodity-specific communities with the aggregate-trade community.

Our results show that commodity-specific community structures are very heterogeneous and in general their statistical properties are quite different from those of the community structure of the aggregate ITN. For example, whereas the number of communities of the aggregate ITN increases in time, this is not the case for most commodity-specific trade networks. Moreover, the shapes and ranges of cluster-size distributions vary a lot across commodities. As far as community structure evolution is concerned, one observes a decrease in concentration over time of cluster-size distributions (a sign of the globalization process) for the aggregate ITN, a pattern that is not always matched at the commodity-specific level, where trade associated to some products has become less and less centralized and increasingly occurs among smaller and more dispersed groups of countries. Furthermore, the community structure of the aggregate ITN has been changing more slowly over time than their commodity-specific counterparts.

We have also explored to what extent the community structure of any commodity-specific network may contribute to shape (or is able to predict) that observed at the aggregate-trade level. We have shown that commodity-specific community structures are becoming more and more similar to the aggregate one, i.e. that the role of all commodities in shaping the aggregated community structure has increased in time. However, a major role for the chemical sector appears from these results, in that they are the ones whose country partition better mimics the aggregate one.

Finally, we have explored two possible factors that correlate with community structure, namely geographical distance and the existence of regional trade agreements between countries. Our findings suggest that geography correlates much more with the observed community structure than RTAs. This result confirms previous findings from the empirical literature on trade.

The paper can be extended and refined in at least three directions. First, our findings related to the impact of geography and RTAs are only partial, as they only check for unconditional effects (i.e. they do not address the residual effects of trade agreements once geography is controlled for). In order to make our statements more robust, one may follow Ref. [20] and compare communities observed in trade data with those detected in the network built with the predictions of a standard gravity model (see for example Ref. [23]). This would allow one to better grasp the role played by country size (GDP) in community-structure formation.

Second, the robustness of our results should be checked against a number of possible issues. For instance, one may consider to apply algorithms allowing for overlapping communities [28,29]. Furthermore, it is well-known that modularity-based community detection suffers from a resolution limit bias [30]. More generally, community-detection approaches based on modularity-function optimization have been shown to suffer, especially in graphs much larger than the ones under analysis here, from problems related to the absence of a clear global maximum [31] and to have lower performance against alternative methods [32]. One of such methods is the community-detection algorithm "Infomap" based on information theory [33,34]. Once applied to our dataset, however, it turns out that only one giant cluster made of all 162 countries is identified in the aggregate-trade data (all years), whereas in the commodity-specific case, cluster numbers and sizes are highly unstable across years and hardly meaningful. As a result, this method delivers quite uninterpretable results. This, we believe, may be due to two reasons. First, as the Authors state, the Infomap method is best suited when one views community structure "as the regularities in the network's topology that allow the greatest compression of the network's structure" [33] rather than, as it happens with modularity maximization, a statistical deviation from a meaningful null model. Second, information-based methods like Infomap are better suited to identify communities in weighted directed graphs when link weights represent flows along paths [34], whereas in the multi-ITN case link weights only proxy bilateral flows (i.e., only a negligible fraction of imported goods are then exported to a third country).

Third, another point that deserves further analysis is the detection of community structures across commodity-specific layers. In the paper, we have analyzed independently the most important 14 layers. This allows one to identify groups of countries that trade the same commodity among them. From an economic point of view this signals strong interdependencies but does not convey any insights on the input–output structure of the cluster. For example, there might be groups of countries that are linked in tightly connected chains or cycles, where a country imports from another one a particular type of commodity needed as input for its peculiar industrial structure, and at the same time exports to other countries in the group another commodity that is fed into their production processes (or consumed as final good). In order to address these issues, one would like to either synthesize into a meaningful statistic all commodity-specific relationships between any two countries or apply new techniques able to detect community structures in multi graphs [35].

Acknowledgements

The authors would like to thank the participants to the NetWorkShop 3.0 (Trento, December 2010) and four anonymous referees for their helpful and insightful comments.

References

[1] X. Li, Yu Y. Jin, G. Chen, Complexity and synchronization of the World Trade Web, Physica A: Statistical Mechanics and its Applications 328 (2003) 287–296.
[2] Á. Serrano, M. Boguñá, Topology of the World Trade Web, Physical Review E 68 (2003) 015101(R).
[3] D. Garlaschelli, M.I. Loffredo, Fitness-dependent topological properties of the World Trade Web, Physical Review Letters 93 (2004) 188701.
[4] D. Garlaschelli, M.I. Loffredo, Structure and evolution of the world trade network, Physica A 355 (2005) 138–144.
[5] A. Serrano, M. Boguñá, A. Vespignani, Patterns of dominant flows in the World Trade Web, Journal of Economic Interaction and Coordination 2 (2007) 111–124.
[6] K. Bhattacharya, G. Mukherjee, J. Sarämaki, K. Kaski, S.S. Manna, The international trade network: weighted network analysis and modeling, Journal of Statistical Mechanics: Theory and Experiment A 2 (2008) P02002.
[7] K. Bhattacharya, G. Mukherjee, S.S. Manna, The international trade network, in: A. Chatterjee, B.K. Chakrabarti (Eds.), Econophysics of Markets and Business Networks, Springer-Verlag, Milan, Italy, 2007.
[8] D. Garlaschelli, T. Di Matteo, T. Aste, G. Caldarelli, M.I. Loffredo, Interplay between topology and dynamics in the World Trade Web, The European Physical Journal B 57 (2007) 1434–6028.
[9] G. Fagiolo, S. Schiavo, J. Reyes, On the topological properties of the World Trade Web: a weighted network analysis, Physica A 387 (2008) 3868–3873.
[10] J. Reyes, S. Schiavo, G. Fagiolo, Assessing the evolution of international economic integration using random-walk betweenness centrality: the cases of East Asia and Latin America, Advances in Complex Systems 11 (2008) 685–702.
[11] G. Fagiolo, S. Schiavo, J. Reyes, World-Trade Web: topological properties, dynamics, and evolution, Physical Review E 79 (2009) 036115.
[12] J. Reichardt, D.R. White, Role models for complex networks, The European Physical Journal B 60 (2007) 217–224.
[13] M. Barigozzi, G. Fagiolo, D. Garlaschelli, Multi-network of international trade: a commodity-specific analysis, Physical Review E 81 (2010) 046104.
[14] C.A. Hidalgo, B. Klinger, A.L. Barabási, R. Hausmann, The product space conditions the development of nations, Science 317 (5837) (2007) 482–487.
[15] C.A. Hidalgo, R. Hausmann, The building blocks of economic complexity, Proceedings of the National Academy of Sciences 106 (26) (2009) 10570–10575.
[16] P.B. Slater, Petroleum trade in 1970: an exploratory analysis, IEEE Transactions on Systems, Man, and Cybernetics (1975) 278–283.
[17] P.B. Slater, W. Schwarz, Global trade patterns: scaling and clustering analyses, IEEE Transactions on Systems, Man, and Cybernetics SMC-9 (7) (1979) 381–387.
[18] S. Fortunato, Community detection in graphs, Physics Reports 486 (2010) 75–174.
[19] I. Tzekina, K. Danthi, D. Rockmore, Evolution of community structure in the World Trade Web, The European Physical Journal B-Condensed Matter 63 (2008) 541–545.
[20] J.A. Reyes, R.B. Wooster, S. Shirrel, Regional trade agreements and the pattern of trade: a networks approach, Working Paper, SSRN Working Paper Series, 2009.
[21] L. Danon, A. Diaz-Guilera, J. Duch, A. Arenas, Comparing community structure identification, Journal of Statistical Mechanics: Theory and Experiment 2005 (09) (2005) P09008.
[22] R.C. Feenstra, R.E. Lipsey, H. Deng, A.C. Ma, H. Mo, World trade flows: 1962–2000, NBER Working Papers 11040, National Bureau of Economic Research, Inc., 2005.

[23] G. Fagiolo, The international-trade network: gravity equations and topological properties, Journal of Economic Interaction and Coordination 5 (2010) 1–25.
[24] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical Review E 69 (2) (2004) 026113.
[25] A. Arenas, J. Duch, A. Fernandez, S. Gomez, Size reduction of complex networks preserving modularity, New Journal of Physics 9 (6) (2007) 176.
[26] F. Glover, M. Laguna, Tabu Search, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
[27] A.K. Rose, Do we really know that the WTO increases trade? American Economic Review 94 (1) (2004) 98–114.
[28] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (2005) 814–818.
[29] V. Nicosia, G. Mangioni, V. Carchiolo, M. Malgeri, Extending the definition of modularity to directed graphs with overlapping communities, Journal of Statistical Mechanics: Theory and Experiment 2009 (03) (2009) P03024.
[30] S. Fortunato, M. Barthélemy, Resolution limit in community detection, Proceedings of the National Academy of Sciences 104 (1) (2007) 36–41.
[31] B.H. Good, Y.-A. de Montjoye, A. Clauset, Performance of modularity maximization in practical contexts, Physical Review E 81 (2010) 046106.
[32] A. Lancichinetti, S. Fortunato, Community detection algorithms: a comparative analysis, Physical Review E 80 (2009) 056117.
[33] M. Rosvall, C.T. Bergstrom, An information-theoretic framework for resolving community structure in complex networks, Proceedings of the National Academy of Sciences 104 (2007) 7327–7331.
[34] M. Rosvall, C.T. Bergstrom, Maps of random walks on complex networks reveal community structure, Proceedings of the National Academy of Sciences 105 (2008) 1118–1123.
[35] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science 328 (5980) (2010) 876–878.