

LETTER • **OPEN ACCESS**

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

To cite this article: S Torreggiani *et al* 2018 *Environ. Res. Lett.* **13** 054026

View the [article online](#) for updates and enhancements.

Environmental Research Letters



LETTER

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

OPEN ACCESS

RECEIVED

15 November 2017

REVISED

16 April 2018

ACCEPTED FOR PUBLICATION

18 April 2018

PUBLISHED

10 May 2018

S Torreggiani^{1,2}, G Mangioni³ , M J Puma^{4,5}  and G Fagiolo¹

¹ Istituto di Economia, Scuola Superiore Sant'Anna, Pisa, Italy

² SOAS University of London, United Kingdom

³ Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, University of Catania, Catania, Italy

⁴ Center for Climate Systems Research and Center for Climate and Life, Columbia University, United States of America; NASA Goddard Institute for Space Studies, NY, United States of America

⁵ Author to whom any correspondence should be addressed.

E-mail: mjp38@columbia.edu

Keywords: complex networks, community-structure detection, multi-layer networks, food security, international trade

Supplementary material for this article is available [online](#)

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Abstract

Achieving international food security requires improved understanding of how international trade networks connect countries around the world through the import-export flows of food commodities. The properties of international food trade networks are still poorly documented, especially from a multi-network perspective. In particular, nothing is known about the multi-network's community structure. Here we find that the individual crop-specific layers of the multi-network have densely connected trading groups, a consistent characteristic over the period 2001–2011. Further, the multi-network is characterized by low variability over this period but with substantial heterogeneity across layers in each year. In particular, the layers are mostly assortative: more-intensively connected countries tend to import from and export to countries that are themselves more connected. We also fit econometric models to identify social, economic and geographic factors explaining the probability that any two countries are co-present in the same community. Our estimates indicate that the probability of country pairs belonging to the same food trade community depends more on geopolitical and economic factors—such as geographical proximity and trade-agreement co-membership—than on country economic size and/or income. These community-structure findings of the multi-network are especially valuable for efforts to understand past and emerging dynamics in the global food system, especially those that examine potential 'shocks' to global food trade.

Introduction

Achieving international food security [1] is undoubtedly one of the major challenges of the forthcoming decades and a globally recognized priority [2]. However, understanding how and why the availability of *and access to* food commodities change across time and space is a dauntingly difficult task, due to its inherent multidimensional nature [3]. International food security may indeed depend on many intertwined phenomena [4], including population growth [5]; agricultural productivity and (over) exploitation of natural resources [6–8]; climate change [9–11];

regional conflicts and epidemics [12]; and the evolution of consumption habits [13–15].

The resulting impact of these interacting factors may generate unexpected volatility and substantial shocks in the supply and availability of food commodities, possibly leading to global crises [16]. International trade, in this respect, may act both as a dampening force and as an amplifying device to regional shocks [17]. On the one hand, international trade may provide new channels to meet increasing food demand through the transfer of food commodities and resources to food-scarce regions. Empirical evidence indeed shows that the amount of traded food has more than

doubled in the last 30 years, and it now accounts for 23% of global production [3]. Furthermore, whereas in the past insufficient domestic production generally implied scarcity in food supplies, production shortfalls in more recent years have been increasingly dealt with by increasing food imports [1, 18].

On the other hand, import-export linkages across countries can boost shock diffusion: increased connectivity in the international trade network (ITN, cf. [19]) can lead to growing fragility [18, 20, 21]. This parallels what happened during the 2007–2008 global financial crisis (GFC henceforth), when seemingly minor shocks spread quickly in a complex, networked world, with disastrous effects [22].

In recent years, a substantial amount of work has been done to explore the network architecture of the aggregate ITN [23]. Furthermore, commodity-specific trade networks have been investigated, both in the case of a set of highly traded commodities, not necessarily food related [24, 25], and for food-trade layers separately [18, 26–32]. However, the multi-network properties of the global food-trade system are still poorly understood [33–35]. In particular, nothing is known about the community structure (CS) of food networks [36], where communities are essentially clusters of vertices characterized by a higher ‘within’ connectivity, but a much sparser connectivity ‘between’ nodes belonging to different clusters.

Lying between the national and global levels, community-level analyses are inherently valuable. They can be considered a proxy for geopolitical relations, which vary depending on the crop of interest [37] and evolve over time. Unfortunately, our progress on quantification of these relations has been limited. In fact, community detection is a very difficult task and a host of different techniques and definitions have been recently proposed in the literature for the case of simple or multi-graphs [36, 38, 39]. Despite the difficulties, identifying communities in a network is fundamental for gaining insights about its structure, its robustness, and the ways in which shocks percolate through it [40]. Indeed, documenting the CS of the international food trade multi-network (IFTMN) may help us better understand how food crises propagate. For example, if trade across countries is organized into well-defined clusters, shocks originating within a cluster might spread more readily within that group than across groups.

Here we start to fill this gap using data on international trade flows taken from FAOSTAT, with a focus on the 16 most internationally traded staple food commodities for the period 1992–2011. We document the evolution of CSs in the IFTMN both across layers (i.e. when the IFTMN is analyzed as a collection of separate layers, each one representing bilateral trade for a specific food commodity, e.g. wheat) and in the multi-layer graph (i.e. when the IFTMN is conceived as a single multi-layer network where countries are connected by multiple import-export relationships, e.g.

Table 1. Top world 16 import commodities in 2013 according to kcal embodied.

Code	Commodity	kcal ^a	USD	% kcal
1	Wheat	6.45×10^{14}	9.71×10^{10}	21.11
2	Soybeans	5.93×10^{14}	1.07×10^{11}	19.43
3	Maize	4.44×10^{14}	4.22×10^{10}	14.54
4	Sugar	2.25×10^{14}	3.31×10^{10}	7.38
5	Rice	1.36×10^{14}	2.61×10^{10}	4.47
6	Barley	1.32×10^{14}	2.74×10^{10}	4.33
7	Oil, Palm	9.74×10^{13}	4.20×10^{10}	3.18
8	Oil, Sunflower	7.22×10^{13}	1.01×10^{10}	2.37
9	Milk	6.81×10^{13}	8.23×10^{10}	2.21
10	Cassava	5.33×10^{13}	4.07×10^9	1.75
11	Pulses	4.64×10^{13}	1.02×10^{10}	1.49
12	Cocoa	4.51×10^{13}	4.22×10^{10}	1.46
13	Pig Meat	4.47×10^{13}	4.21×10^{10}	1.43
14	Poultry Meat	2.82×10^{13}	3.45×10^{10}	0.92
15	Nuts	2.61×10^{13}	2.03×10^{10}	0.86
16	Sorghum	2.40×10^{13}	2.01×10^9	0.78

^a Source: Our computation on FAOSTAT data (see fao.org/faostat).

for maize, wheat, rice, etc.). We then fit econometric models to identify social, economic and geographic factors explaining the probability that any two country are co-present in the same community.

Materials and methods

Data and definitions

We use FAOSTAT data on international trade flows, which contain bilateral export-import yearly figures for food and agricultural products in the period 1986–2013. (Data are available at fao.org/faostat.) Of the products available, we select the 16 most-traded commodities in 2013, ranked according to the total kilocalories (kcal henceforth) embodied, so as to account for about 90% of the total kcal trade for food-related goods. To compute total kcal embodied we explicitly consider caloric values of secondary and derivative products, see table B1 in appendix B for details available at stacks.iop.org/ERL/13/054026/mmedia. Primary and secondary products are aggregated after converting them to kcal.

Table 1 lists the top 16 commodities according to kcal embodied (in 2013) and their trade value (in current USD). As expected, the two rankings are not correlated. For example, there are traded commodities with an extremely high economic value that contribute much less in terms of kcal (e.g. meat and animal products). The most traded products on a value basis are meat; fruits and nuts; and coffee. Notice also that the distribution of kcal is extremely skewed: more than 55% of total kcal are accounted for by wheat, soybean, maize and rice, which together form just 23% of total value in USD.

Selecting commodities according to a mass-to-kcal conversion—rather than value or volume—allows us to aggregate primary crops together with their processed commodities to understand global food trade from a caloric food-security perspective. Other

dimensions of the global food system—such as the economic [41], nutritional [42], or virtual water [32, 43–46] properties—may have distinct characteristics and may be analyzed in future studies.

In order not to bias our analysis with issues related to the collapse of the USSR and of the former Yugoslavia, we do not include the years 1986–1991. We also remove the two most recent years (2012–2013) from the sample, as updated bilateral data are still not available for some products and/or countries⁶. We include a country in our sample if it is involved in a positive bilateral flow for at least one year or one commodity, which gives us $N = 178$ countries (see table A1 in appendix A for a complete list), whose bilateral trade flows for the 16 selected commodities are observed from 1992 to 2011 ($T = 20$).

Network structure

We define the IFTMN as the sequence of T multi-layer networks, where each layer represents bilateral trade among our N countries for a specific commodity $c = 1, \dots, C$ ($C = 16$) in a given year. More formally, in each year $t = 1992, \dots, 2011$, let \mathbf{X}^t be the 3-dimensional weight matrix whose generic entry $x_{ij,c}^t \geq 0$ represents exports (in kcal) from country i to country j for commodity c in year t . As usual, we posit that $x_{ii,c}^t = 0$ for all i, c and t . We define the IFTMN as the time sequence of multi-layer networks characterized by the time sequence of weighted-directed matrices $\{\mathbf{X}^t, t = 1, \dots, T\}$. In other words, each snapshot (year) of the IFTMN is a multi-layer network, where the nodes are the 178 countries connected by multiple directed links (or edges), each of which represents an exporter-importer flow for a particular commodity, weighted by its correspondent intensity in terms of kcal traded. A directed link $(i \rightarrow j)_c^t$ is therefore present for a given commodity-year combination (c, t) if i exports to j a non-zero volume for commodity c in year t . All zero off-diagonal entries therefore represent either a missing value or a sheer zero-trade flow.⁷

Prior to performing community detection, we explore the properties of the time sequence of multi-networks \mathbf{X}^t using a principal component analysis in the space of network statistics computed over each single layer. More precisely, given link weights $x_{ij,c}^t$ of layer (c, t) , let \mathbf{W}_c^t be the associated log-transformed weight matrix.⁸ Also, we define \mathbf{A}_c^t as the correspondent adjacency matrix, which is the $N \times N$ binary matrix whose generic element a_{ij} equals

one if there exists a link from i to j and equals zero otherwise.

In each year t , we compute a number of network statistics to fully characterize the weighted and binary topological properties of the layer. These network statistics are computed over the weight \mathbf{W}_c^t and adjacency \mathbf{A}_c^t matrices of each layer (c, t) , respectively. These metrics include: (i) *density*, defined as the existence number of links over all possible $N(N-1)$ directed edges; (ii) *bilateral density*, defined as the ratio of reciprocated links; (iii) *weighted asymmetry* as defined in [47]; (iv) *size of largest connected component* (LCC), defined as the number of nodes in the largest connected subgraph, where connectivity is defined in a weak form (i.e. disregarding directionality); (v) *centralization*, see [23], which measures how much the binary structure is centralized; (vi) *binary/weighted assortativity*, defined as the correlation coefficient between node average nearest-neighbor degree/strength (ANND/S) and total node degree/strength, see [23]; (v) *binary/weighted average clustering*, defined as the average across nodes of node total binary/weighted clustering coefficients (see [49]); and (vi) *average and standard deviation of link weights*, defined as the arithmetic average and standard deviation of the log-transformed export flows in a single layer.

We note that, whereas bilateral density measures symmetry at a binary level, the weighted-asymmetry index employs link weights to assess how much reciprocity is present in the weighted directed graph. Also, the assortativity metrics are to assess the tendency of nodes to connect to other nodes with similar properties. If the indices are positive, the graph is assortative, meaning that nodes tend to connect to other nodes with similar properties; conversely, negative indices indicate that nodes tend to connect to those with dissimilar properties (i.e. they are disassortative).

This set of 11 metrics can be used to provide insight into the topological characteristics and potential vulnerabilities of each layer. However, many of these metrics are closely related and are possibly redundant (i.e. too highly correlated with the most basic statistics like density). We therefore perform a principal-component analysis to reduce the dimensionality of the space of remaining statistics and then interpret the results. This allows us both to identify network measures that better characterize the topological structure IFTMN in each year and to explore similarities and differences among commodity networks.

Community structure detection

Here, we tackle the problem of community detection by treating the IFTMN as a collection of C different commodity-specific weighted-directed simple graphs in any given t and analyzing the CS of each layer separately. To identify communities, we employ the modularity optimization approach originally introduced by [50] and subsequently extended to the case

⁶ Note that our selected commodities are still the top-16 most-traded agricultural products in terms of kcal also in 2011.

⁷ In the IFTMN, links between any two commodity layers c_1 and c_2 , $c_1 \neq c_2$ are present only between *copies* of the same country, i.e. any country i is connected to itself in all the layers. Two different countries are not linked across different layers. In this respect, the IFTMN can be defined as a *multiplex* or *colored network*.

⁸ As it is customary in this literature [19], positive trade levels are log-transformed in order to reduce the skewness of their distribution.

of weighted directed graphs by [51]. In this case, the modularity function to be maximized is:

$$Q_c^t = \frac{1}{X_c^t} \sum_{ij} (x_{ij,c}^t - E[x_{ij,c}^t]) \delta(\xi_{i,c}^t, \xi_{j,c}^t), \quad (1)$$

where X_c^t is the volume of the layer (c,t) and δ is a Kronecker delta function equal to 1 if nodes i and j are in the same community and 0 otherwise. E is the expected value of the link weight $x_{ij,c}^t$, which following [51] reads:

$$E[x_{ij,c}^t] = \frac{s_{i,c}^t(out) \cdot s_{j,c}^t(in)}{X_c^t}, \quad (2)$$

where $s_{i,c}^t(out)$ and $s_{j,c}^t(in)$ are respectively out-strength of node i (i.e. the sum of outward link weights) and in-strength of node j (i.e. the sum of inward link weights) [52]. To optimize Q_c^t , we employ the modularity-clustering heuristic developed by [53], which extends and improves the well-known ‘Louvain’ algorithm pioneered by [54] (see appendix C for more details). This procedure ends up, for any given year t and commodity-layer c , with a univocal assignment of countries into clusters, the number of which is not fixed ex-ante, in such a way that each country belongs to a single cluster (i.e. communities are not overlapping). Clusters can also contain a single country, e.g. if that country is an isolated node in the network. Note that we check the results of the above procedure by treating the IFTMN for any t as a single multi-layer network (see appendix C for further details).

Econometric models

As mentioned, identifying communities in the IFTMN treated as a collection of C separate layers, results in a univocal assignment of countries to clusters for any given choice of t and c . Clusters are multilateral entities, as they emerge whenever a group of countries trades comparatively more among them than they do with countries outside the cluster. But what are the factors underlying the emergence of such clusters? Here, we address this issue fitting probit and logit models [55] that explain the probability that any two countries belong to the same cluster (for a given (c,t) slice of the IFTMN) as a function of economic, socio-political and geographical, bilateral relationships. More precisely, we perform two sets of exercises.

First, for all $c=1,\dots,16$ and two selected years ($t_0=2001$ and $t_1=2011$)⁹, we fit to the data the following probit model using a maximum-likelihood procedure:

$$Prob\{\gamma_{ij,c}^t = 1\} = \Phi(\alpha + \beta Z_{ij}^t), \quad (3)$$

where $\gamma_{ij,c}^t$ is a binary indicator for the event that countries i and j belong to the same community for product

c and year $t \in \{t_0, t_1\}$, Φ is the cumulative distribution function for the standard normal variate¹⁰, α is a constant, β is a vector of slopes and Z_{ij}^t is a set of bilateral covariates (more details below).

Second, we run a panel-data estimation of the probit model in equation (3) on the pooled dataset containing all the years in our sample, for some selected commodities (i.e. wheat, maize and rice). We choose wheat, maize, and rice (and their associated commodities) as they are among the most important internationally traded grains and are fundamental to staple food supplies around the world. Panel estimations feature the same covariates of the cross-section setup, but they now become time-varying. Furthermore, as it is customary in this approach [56], we control for unobserved heterogeneity and common trend effects including in panel regressions both time-invariant country fixed-effects and time dummies.

To choose the covariates, we rely on the literature on the empirical trade-gravity model [57], see appendix E and table E1 for details. We employ five classes of covariates: (1) economic variables (i.e. combined measures of economic country size and income); (2) trade policy variables (e.g. whether the two countries belong to the same preferential trade agreements); (3) geographical variables (e.g. distance between countries and whether they share a border); (4) historical/political variables (e.g. former colonial relationships); and (5) cultural variables (i.e. whether countries share the same language).

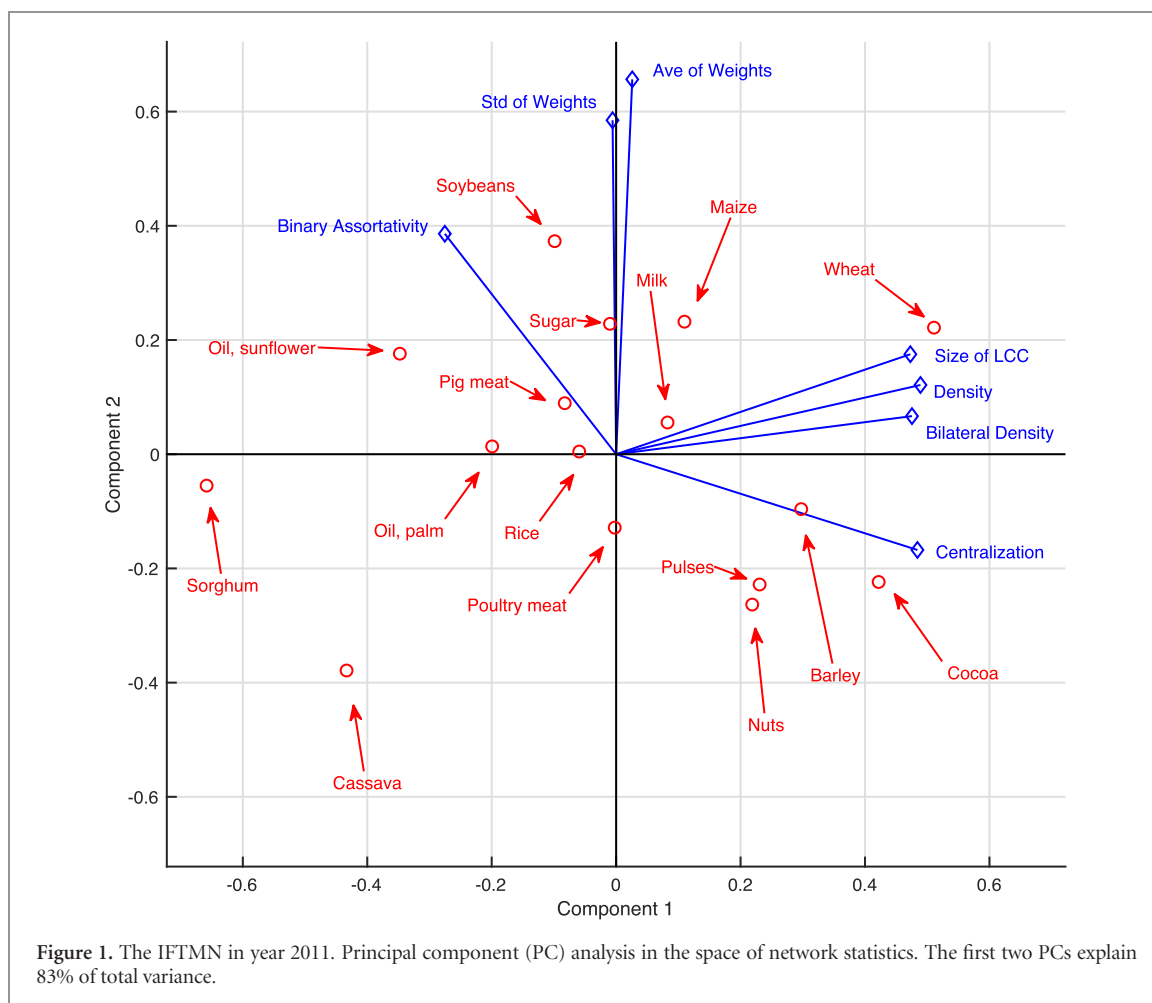
Despite the fact that our probit specification has an obvious gravity flavor, it departs from traditional trade-gravity models in the way we treat directionality of relationships. Indeed, because the co-presence relations are symmetric by definition, the binary response model in equation 3 does not distinguish between importer and exporter, as, on the contrary, gravity models with trade flows as dependent variable often do. Therefore, sign and intensity of the impact of covariates cannot differ between origin and destination markets.

Results

We now turn to a description of our main results. First, we describe some basic network properties of the IFTMN, both across commodity-layers and time. Second, we discuss the CS of ITMN considered as a collection of C separate layers. Third, we explain co-presence in clusters using probit models. Finally, we check what happens when CS detection is performed over the IFTMN described as a multi-layer network.

⁹ These two years have been chosen in order to focus on two time periods sufficiently far from the GFC.

¹⁰ All our econometric results are robust when we employ a logit specification instead of a probit, i.e. when we let ϕ be the cumulative distribution of a logistic random variate.



Overview of network properties

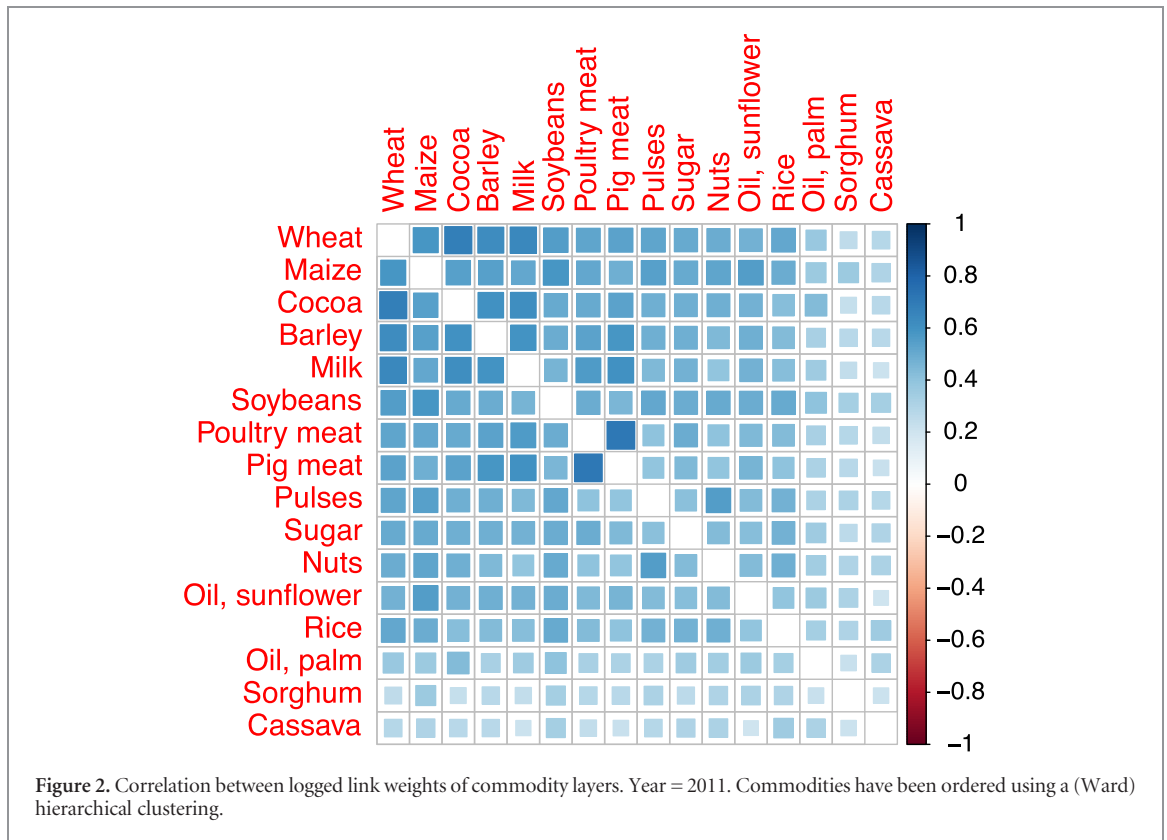
The IFTMN is characterized by substantial heterogeneity across commodities (i.e. layers) but low variability over the time interval under observation. A comparison of results in tables F1–F2 in appendix F, which report network statistics in 2001 and 2011, suggests that network structure did not go through dramatic changes before and after the GFC.

However, our analysis indicates considerable variation in the topological properties across commodity layers. For example, the IFTMN is composed of small-density layers (as compared to the aggregate ITN), whose link probabilities range from 0.01–0.16. Substantial variation is also detected in the size of the largest connected component (LCC)—from 87 to 171—and many other statistics. Therefore, a principal component (PC) analysis can help summarize the most important dimensions of variability. Results for the year 2011 are reported in figure 1. We use a bi-plot to represent both the units (i.e. commodities) in the space of the first two PCs (which together explain 83% of total variance) and network statistics as vectors (whose direction and length indicate how each variable contributes to the two principal components in the plot).

Starting with the network statistics (in blue), the first PC is positively correlated with connectivity measures (i.e. density and size of LCC), network

symmetry, and centralization and is negatively correlated with binary assortativity (i.e. the larger the x -axis coordinate, the smaller the assortativity coefficient). The second PC is instead positively correlated with average and standard deviation of link weights (as well as assortativity). This means that, overall, commodity layers tend to have higher density and LCC size and to be more centralized and symmetric but, at the same time, less assortative. Additionally, more intense bilateral connections are gained, on average, at the expense of a larger standard deviation thereof.

Next, we consider the PC analysis for the commodities (in red). The position of layers in the bi-plot suggests the existence of two paradigmatic cases. The first one is represented by layers such as wheat, cocoa, and barley, which are characterized by relatively high connectivity, centralization, and symmetry but a relatively smaller assortativity and a lower intensity and variability of import-export relationships. To the second one belong layers such as sorghum and cassava, which are much less connected and symmetric, and they are structured over more intense and less variable trade relationships. Other important layers like maize, rice and soybeans play instead an intermediate role, being less internally connected than wheat but displaying stronger and more variable bilateral connections.



Network statistics in tables F1–F2 and their correlations (see figure F1) reveal two important additional facts. First, the layers of the IFTMN are mostly assortative: more-intensively connected countries tend to import from and export to countries which are themselves more connected.

Second, the weighted version of statistics such as asymmetry, clustering and assortativity are almost linearly correlated with their binary counterpart, suggesting that in the IFTMN, unlike in the aggregate ITN, the creation of new trade channels are more important than increases in trade flows of already existing connections (i.e. in economics jargon, extensive trade margins are more important than intensive ones).

We now explore across-layer correlation in (logs of) link-weight distributions $w_{ij,c}^l = \log(x_{ij,c}^l)$, cf. figure 2 for year 2011 and figure F3 in appendix F for year 2001. We notice that almost all commodities are traded as complements (i.e. all correlations are positive and significant). The only exceptions are palm oil, sorghum and cassava, which are traded in an almost uncorrelated way with all the others. This may probably be due to the fact that these are either markets extremely concentrated around a handful of producers (i.e. palm oil) or extremely agglomerated geographically (i.e. cassava and sorghum).

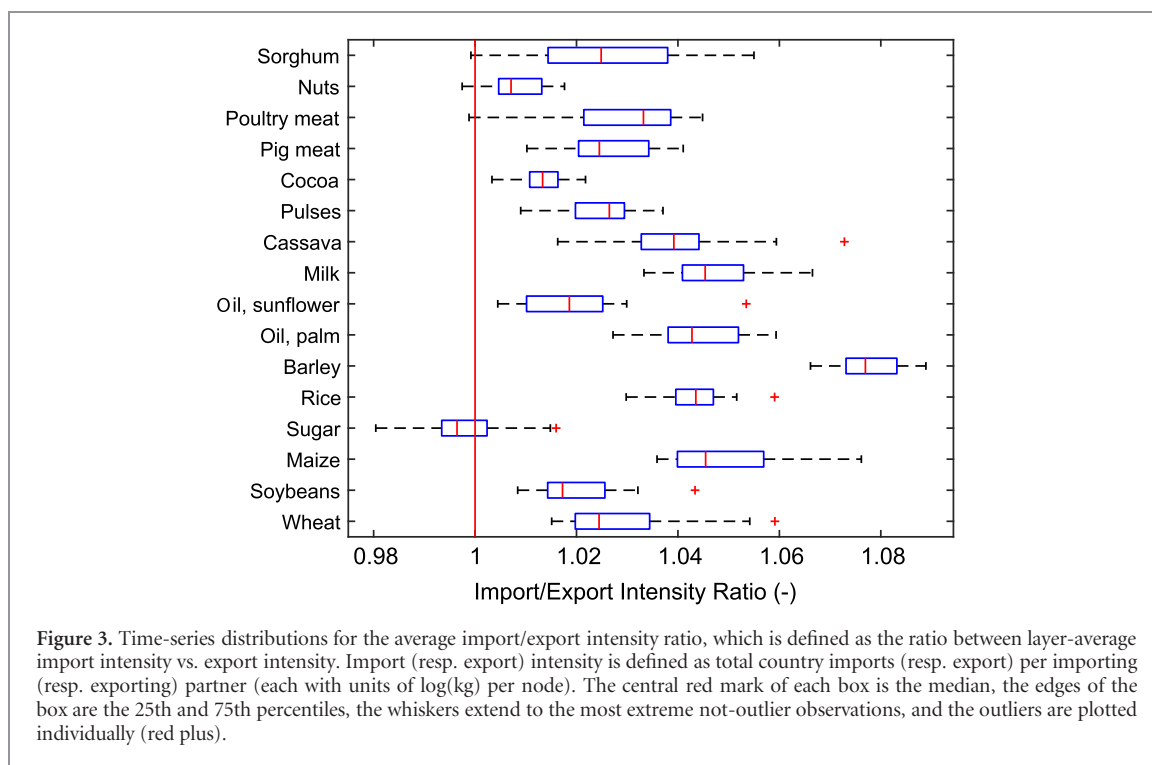
Finally, we investigate the extent to which export per outward link is associated with imports per inward link, across years and layers. Figure 3 depicts time-series distributions for the ratio between layer-average import intensity vs. export intensity (i.e. the import/export intensity ratio). Import intensity is defined as total

country imports per importing partner (in network-science jargon, the ratio between node in-strength and node in-degree). Likewise, export intensity is defined as total country exports per exporting partner (i.e. the ratio between node out-strength and node out-degree). Note how almost all layers have been characterized by ratios always larger than one across the years. This means that, on average, countries tend to have—irrespective of the commodity traded and its share on the world market—more intensive import relations than export ones. This result is consistent with the evidence shown by [24] for a more aggregated set of commodity-specific—not necessarily food-related—networks (and it is, in particular, true for coarse cereals). This evidence could be a symptom of the high dependency of several countries on a small number (say one or two) import channels for their staple-food supply.

Layer-by-layer community structure

We now discuss community-detection findings when the IFTMN is treated, in each year, as a collection of independent food-staple trade layers. We begin with results related to two temporal cross sections—for the individual years 2001 and 2011—across all layers. Then, for three selected commodities (wheat, maize and rice), we document the evidence on community-detection for the 2001–2011 panel.

As table I1 shows, the first general observation is that the IFTMN exhibits a very high level of (maximum) modularity in almost all layers and years. This suggests that the IFTMN is characterized through-



out by a strong community structure, with countries that organize into densely linked groups. Indeed, maximum modularity levels typically fall in the range [0.2,0.5], which, as suggested in [50], is strong evidence for the existence of well-defined clusters. The only exception to this general rule is cassava, which displays an almost negligible level of modularity. In each layer, we identify on average 6 clusters (or communities) with number ranging from 3 (for poultry meat in 2011, the least dispersed layer on average) to 10 (for sorghum in 2001, the most dispersed layer on average).

More importantly, our community detection exercises indicate that countries in the IFTMN tend to cluster into trading blocs that display relevant geopolitical and socioeconomic patterns. This can be seen in figure 4, where we plot choropleth maps with countries colored according to their community membership in 2011 for selected commodities.

Choropleth maps for year 2011 reveal interesting cross-layer regularities. First, there often exists a North American cluster (with the US and Canada often linked to Central and Latin America countries), whereas relevant breadbaskets such as Brazil and Argentina often set up alternative communities independently. Second, Russia generally forms a cluster together with Central, Caucasian and East-European (non EU-members) states, often absorbing some MENA region countries (especially Egypt). A unified European cluster often emerges, sometimes linked with the Russian cluster and rarely linked with the US, confirming that Europe is not such an open market for many agricultural products. Furthermore, a consolidated and independent Asian cluster seems to exist only in the

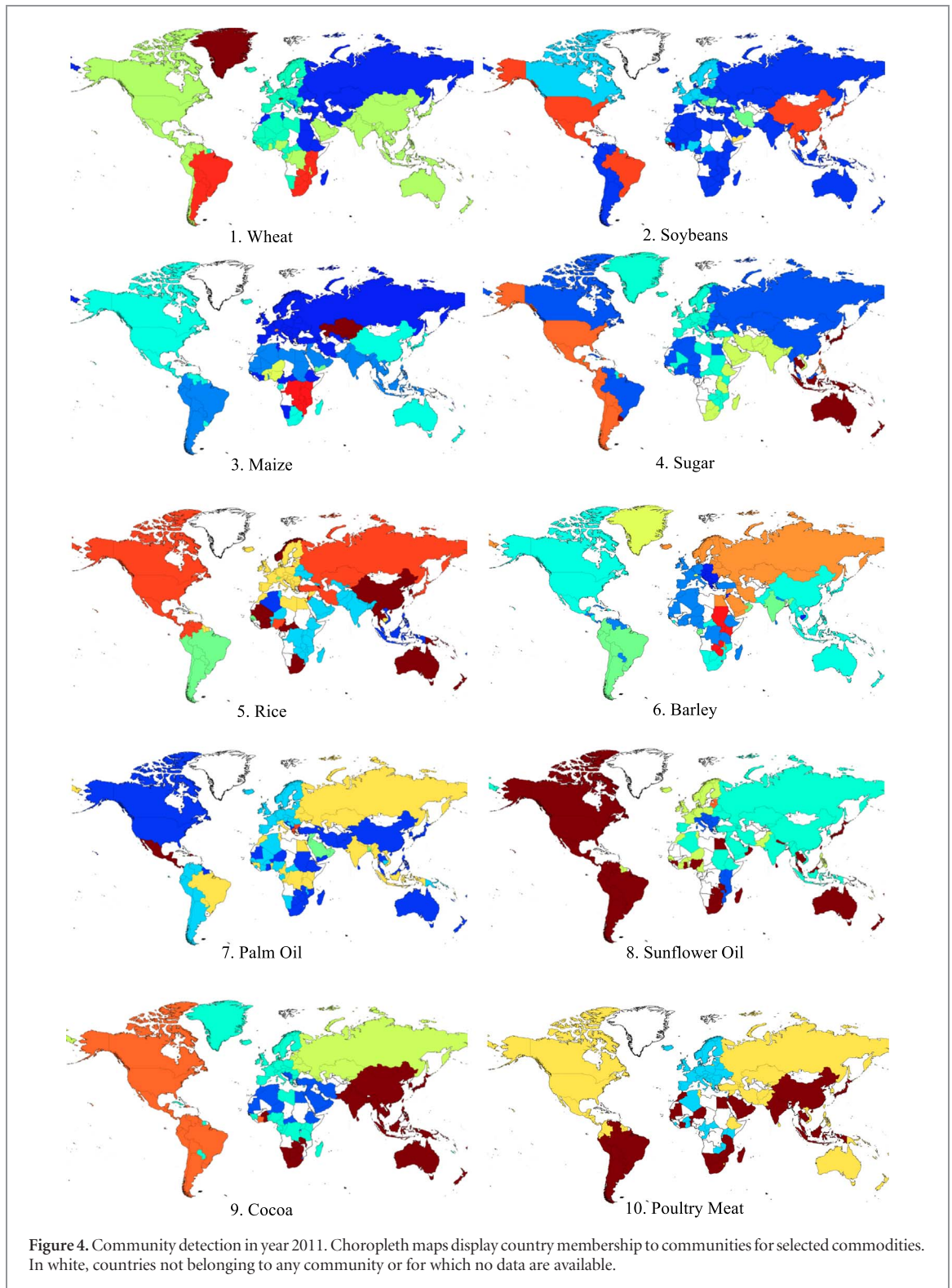
case the region is a net importer for that commodity (i.e. wheat, milk and dairy products, and cocoa). East Asian (e.g. China, India and Japan) and Southeast Asian (e.g. Vietnam, the Philippines, and Thailand) countries instead typically belong to different communities, orbiting around other clusters such as the North American and South American ones. Finally, Africa and the Middle East are often divided—independently of the commodity examined—and only in a few cases we can observe a small independent Eastern Sub-Saharan cluster.

Apart from these macro regularities, several cross-sectional differences also emerge among commodity-specific community structures¹¹, the most striking of which concerns concentration in their size distributions (see figure 5 in appendix D for the case of year 2011). The most concentrated community structures are those of soybeans, palm oil, poultry meat and nuts, whereas rice exhibits the most homogeneous size distribution.¹²

Similarities and differences among community structures can be better appreciated computing the normalized mutual information (NMI) index between pairs of community structures (see figure 5 and appendix D for details). The NMI index ranges between 0 and 1 and increases the more the two community structures are similar. Three groups of commodities can be identified (outlined by the three squares

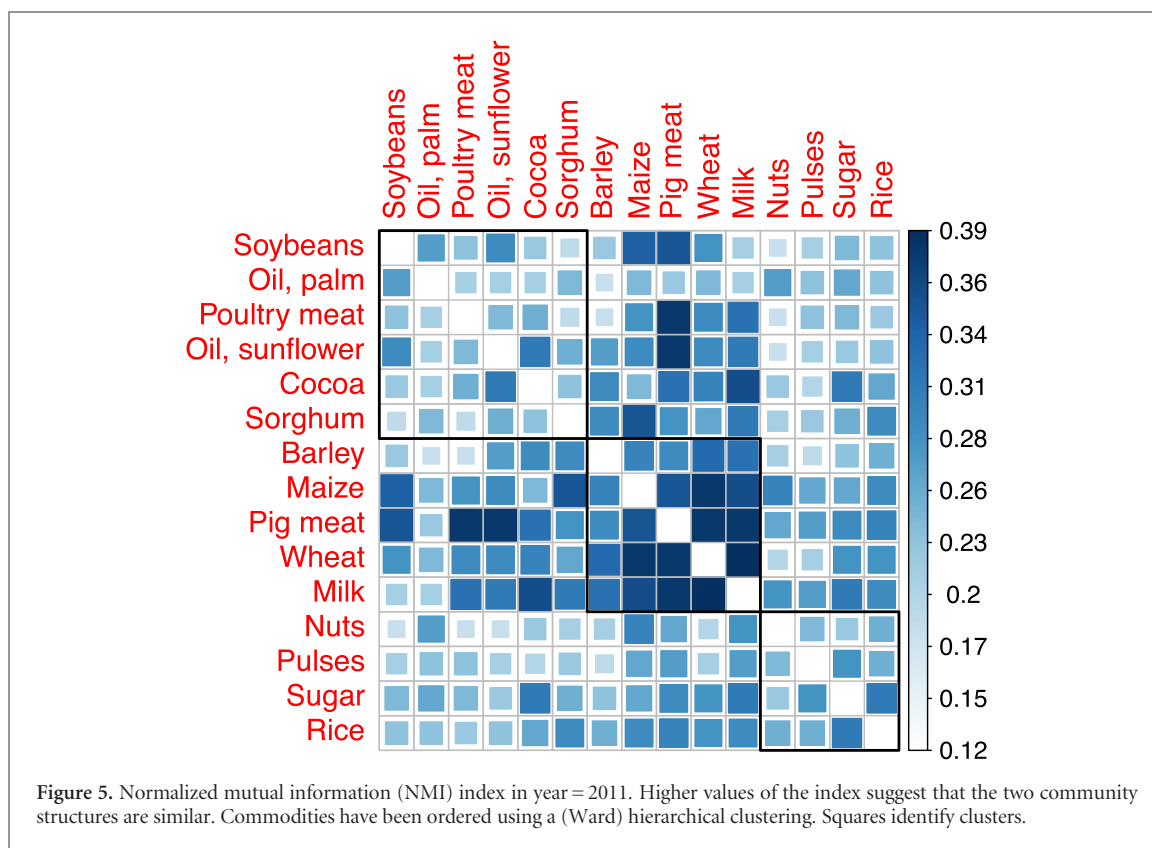
¹¹ In appendix G we discuss in details economic factors that can explain the pattern of each commodity-specific community structure in 2011.

¹² This result is confirmed when one computes the Herfindahl concentration index (see description that follows).



in the figure). The first one comprises the most similar structures, i.e. coarse grains (barley, maize, wheat), pig meat and milk. The other two consist of commodities that exhibit quite different trading blocs, and differ from the other groups. These are: (i) nuts, pulses, sugar and rice; and (ii) soybeans, poultry meat, oil, cocoa and sorghum. Note that pig and poultry meat are very similar in terms of their community structures but consist of different groups.

We now explore whether community structures have changed from 2001–2011. Figure I1 in appendix I shows, for a few commodities, country community membership in 2001. A qualitative comparison with figure 4 shows that in 2011 the European trading bloc became larger, possibly due the Eastern enlargement of the Union (from 15–27 members). The evidence is particularly strong in the case of wheat, maize, sugar, rice, palm oil and cocoa, whereas the finding holds



to a lesser extent for barley, milk, pulses and poultry meat. Overall, this expansion may be interpreted as evidence of the effectiveness of the Common Agricultural Policy (CAP) of the European Union. Furthermore, comparing 2001 and 2011 maps reveals an increasing influence of Brazil, Russia, India and China (i.e. the BRIC countries) in the African continent. This evidence may be partly explained by the increasing hegemony of Russia and India in Eastern Africa, which has gradually undermined that of Australia in wheat and rice trade. Similarly, maps seem to be coherent with the increasing importance that Brazil gained as maize supplier in African and Middle Eastern countries, at the expense of the North American and the European clusters. (Additional description of community structures can be found in appendix G.)

More generally, community structures in 2001 differ from those in 2011, because the size distributions of the latter are typically more concentrated. Figure I3 in appendix I plots the normalized Herfindahl concentration index (H index) computed in 2001 and 2011 for all commodity networks (except cassava) and shows that the lion's share of layers lie above the main diagonal. Rice, soybeans, poultry meat and sunflower oil display the largest increase in concentration. A more concentrated community structure implies that a larger share of countries belong to existing trading groups. Therefore, increases in H index can be interpreted as a tendency to a more globalized trade network. Notice that increasing concentration levels are not necessary associated with a decrease in the number of detected communities (cf table I1). This suggests

that, when detected, increasing concentration levels in community size distributions are attained through country switching among clusters and not due to a reduction in the number of trading blocs.

To delve further into the time dynamics of community structures, we focus on three selected commodities: wheat, maize and rice. We document how community structure for these three commodities evolve across the whole time sample (1992–2011). Figure I4 presents the time series of community number (left) and maximum modularity (right). Note that, in general, modularity has been increasing over time, suggesting that the IFTMN, at least in the three layers considered in the figure, has exhibited a stronger and stronger tendency to clusterize into well-defined trading blocs. Furthermore, the three commodities considered have followed quite distinct time patterns as far as the number of detected communities is concerned. Maize trade network has been organizing itself into an increasing number of clusters, whereas the number of trading blocs in the wheat network has decreased and stabilized around four. Finally, the rice network has experienced turbulence, oscillating between six and nine trading groups over time.

Econometric models

As visual inspection of figure 4 for 2011 and figure I1 for 2001 shows, community structures in the IFTMN exhibit clear geopolitical and socioeconomic regularities. In order to quantitatively explore this issue, we run a set of probit-regression exercises to examine

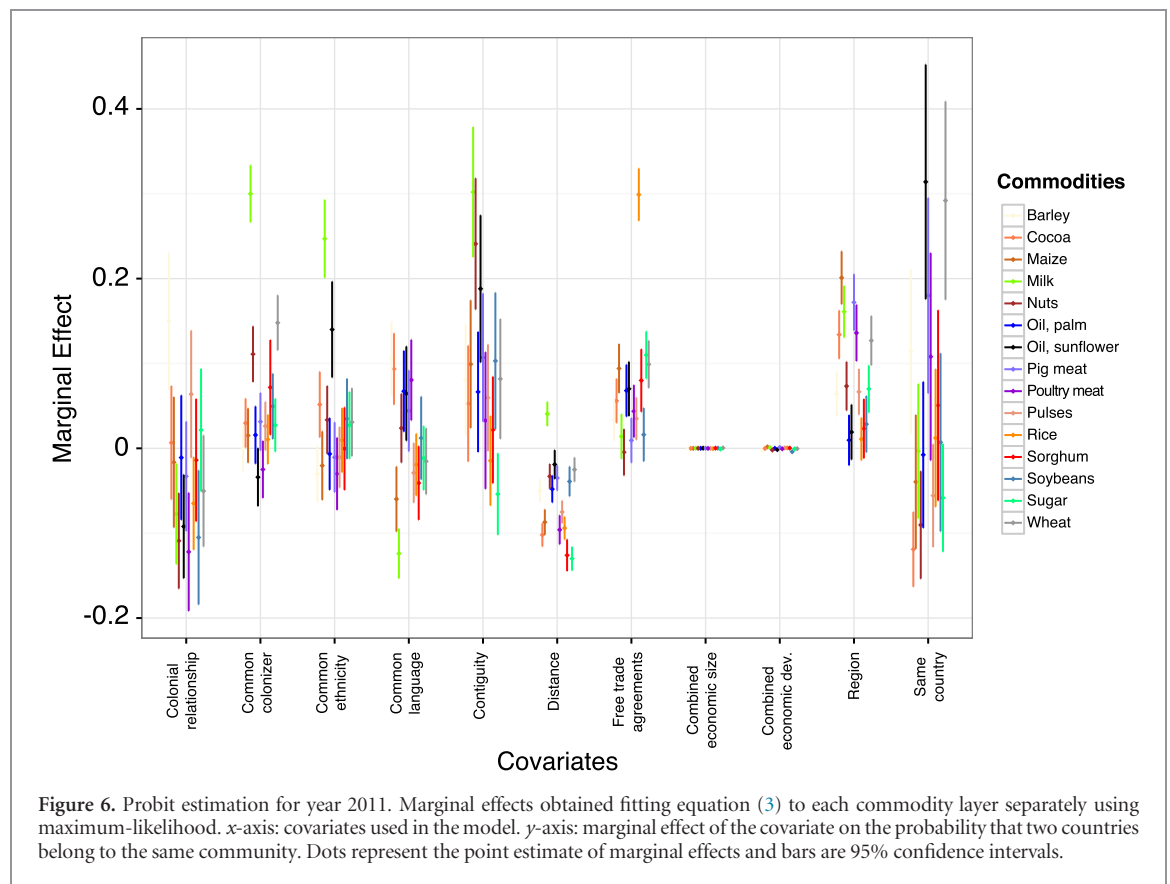


Figure 6. Probit estimation for year 2011. Marginal effects obtained fitting equation (3) to each commodity layer separately using maximum-likelihood. x-axis: covariates used in the model. y-axis: marginal effect of the covariate on the probability that two countries belong to the same community. Dots represent the point estimate of marginal effects and bars are 95% confidence intervals.

the probability that any two countries belong to the same trade bloc as a function of a host of covariates (see section 3.3 and table E1), capturing country-pair (dis)similarity along geographical, economic, social, and political dimensions.

Covariates employed in the analysis are borrowed from the trade-gravity literature [57], which suggests that bilateral trade flows typically increase with importer and exporter market size and income (proxied by country total and per-capita GDP) and decrease with larger trade friction. The latter is usually proxied by geographical distance and a number of bilateral indicators (e.g. dummy variables) that control—among other things—for whether the importer and the exporter share a border, a common language, a trade agreement, any colonial relationship, and whether they belong to the same geographical macro-area.

We begin by fitting equation (3) cross-sectionally to year 2001 and year 2011, for all commodity layers. Results for year 2011 are visually presented in figure 6, where point estimates of marginal effects of covariates are plotted together with their 95% confidence intervals for all commodities (see figure I5 in appendix I for year 2001)¹³.

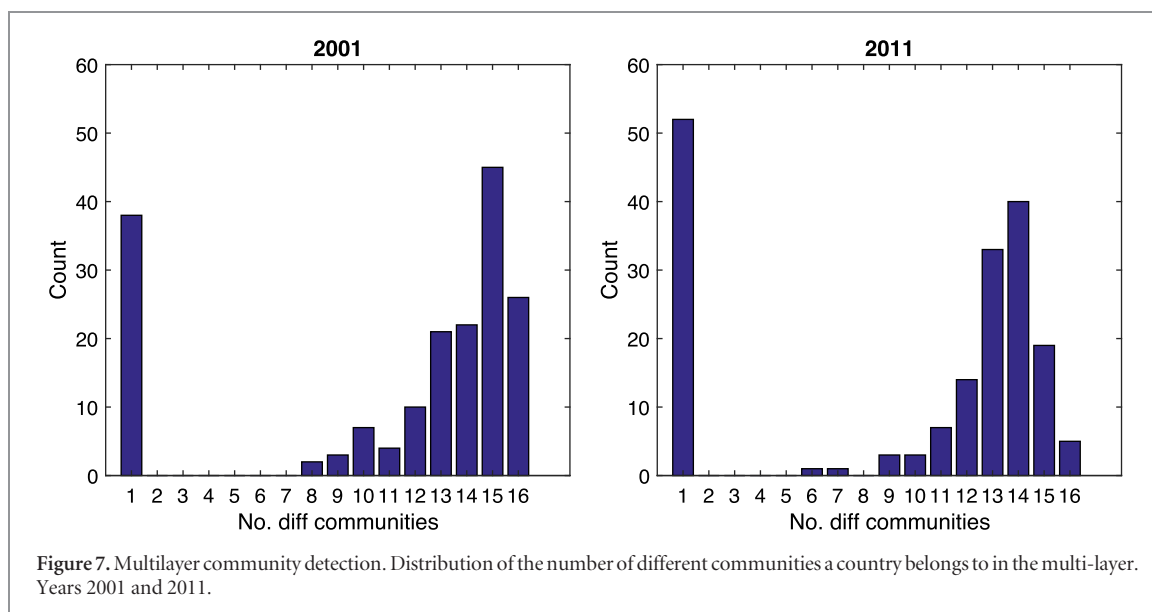
Our findings indicate that distance has a negative and statistically significant impact on the probability

that two countries belong to the same trade community, for all products considered except milk. Other geographically-related covariates (such as contiguity and regional membership) have a product-specific effect, both in terms of significance and sign; nevertheless, they generally boost the co-presence of country pairs in the same trade bloc. Furthermore, free-trade agreements almost always promote co-presence, and their importance has become higher in 2011 as compared to 2001. The role of past colonial relationships and common language is less relevant in explaining joint membership. Most importantly, regressions suggest that economic indicators (i.e. absolute and per-capita GDP) are not significant either in statistical or economic terms, because of high standard errors and small marginal effects.

These results are confirmed by panel-data exercises run for the cases of wheat, rice and maize. We regress co-presence probabilities against the same set of covariates used in the cross-section setup, but now employ the entire time sample in a dynamic fashion and control for common trends and country-specific unobserved heterogeneity with an appropriate use of dummy variables. Again, as figure 16 shows, distance and free trade agreements¹⁴ are two important determinants of the co-presence of country pairs in the same

¹³ All models turn out to be nicely specified according to standard goodness-of-fit tests, e.g. the Akaike information criterion (AIC).

¹⁴ More precisely, the EU27 trade agreement and NAFTA seem to strongly affect co-presence probabilities, as well as AFTA for maize and EFTA for wheat.



trade community, whereas economic factors are generally weak or insignificant.

Overall, our econometric estimates are in line with the trade-gravity literature, as they show that distance, trade frictions and trade agreements are important determinants of country co-presence in trade communities as they are for bilateral trade flows. However, they strongly depart from traditional gravity exercises as they indicate a very weak impact of country economic size and income in shaping food-trade blocs, whereas it is well known that these two covariates explain to a great extent the intensive margins of aggregate trade [57]. (see appendix H for additional discussion.)

Multi-layer community detection

In the last subsection, we have performed a community-detection analysis assuming that the IFTMN consists of independent layers in each time period. Here, we explore how communities would look if they could span across layers. More precisely, we suppose that each country is coupled with itself across commodity slices. Therefore, in each year, the IFTMN becomes a multi-layer network, where nodes are country-commodity pairs. Identifying communities in such an object means finding clusters where countries and commodities can possibly repeat themselves many times: the same country (respectively, commodity) may belong to different clusters as it can appear coupled with different commodities (respectively, countries).

We use this analysis to explore the shape of clusters in the multi-network. To do so, we begin by studying the distribution of the number of different communities a country belongs to, which we interpret as a rough measure of country diversification in the IFTMN. The intuition is that a country belonging to a small number of different communities tends to be mostly connected with instances of ‘itself’ in different

commodity layers and therefore depends on the same group of other country-commodity pairs for all possible staple-food products it trades. Conversely, if a country appears in a large number of different communities in the multi-network (and thus is never isolated) then it relies on several different clusters of country-product pairs depending on the specific product it trades.

As we show in figure 7, the frequency distribution of this statistics are markedly bi-modal, with a peak at 1 and another peak around 14–15. This suggests that community structures in the multi-layer are polarized into two groups. The first one consists of countries that—irrespective of the commodity traded—always belong to the same community in the multilayer. These are countries that are poorly diversified and are the least networked in the food-trade system. Countries in the second group belong instead to several different communities depending on the commodity traded and therefore are highly diversified in the multilayer. This finding is relevant for food-security issues as it suggests that countries belonging to the first group may be more vulnerable than those in the second group to shocks that put at risk the supply of one or more food commodities. The geographical distribution of the two groups of countries is depicted in figure 8. Notice how the first group is mostly located in Africa but also features countries in the Middle East and Asia.

Discussion and conclusions

The topology of the international food trade multi-network is key to understanding vulnerabilities in the global food system. We show that the IFTMN is increasingly globalized and characterized by substantial heterogeneity across commodities. These findings highlight the need to account for each commodity

network's unique properties when considering food policies [37], especially as the multi-network changes over time. Another key finding is that countries tend to have more intensive import relations than export ones. From a food-policy perspective, it highlights the importance of understanding trade dependences and their link with robustness or vulnerability of a country.

Our analyses also show that the individual layers of the IFTMN have densely connected trading groups, a consistent characteristic over the period 1994–2011. At the same time, we show that these trading groups are evolving. For example, we present evidence that the European trading block increased in size and that BRIC countries have expanded their influence in Africa. In addition, the rice network has experienced significant turbulence (relative to the other commodity layers) in terms of community structure, which is important given the prominence of rice in traditional Asian diets. We also uncover important geographic features. For example, East Asian and Southeast Asian countries typically belong to different communities, orbiting around other clusters, while Africa and the Middle East are often divided in terms of community membership.

Our community structure findings are important, as they fundamentally affects how a shock would spread within the global food system. If, for example, the epicenter of a shock is within a community, we would expect that countries in this community would face a two-fold challenge: (1) reduced supply from domestic production and/or from their usual import partners and (2) high international prices. To the extent possible, governments and companies within these countries would adjust their procurement strategies to find new sources from members of the other trading communities. Outside of the epicenter community, network characteristics like inter-community connectivity and other global dynamics like trade interventions would be critically important.

One straightforward application of the knowledge generated from understanding commodity specific community structures is that we can improve our understanding of potential vulnerabilities to various disruption scenarios. First let us consider a major disruption to rice production. In a scenario where China experiences a major negative production shock, how would the community structure of the rice network modulate global impacts? China would look to the international markets to make up for any shortfall that its food reserve system could not handle. Four of the top five exporters—Thailand, Vietnam, India and Pakistan—are co-located in Asia, where Thailand is in the same community as China, Vietnam is part of a predominantly Southeast Asian community, and India and Pakistan are both in another community. Therefore, the burden of making up for the Chinese production shortfall would fall primarily on Asian countries, with perhaps the US also contributing

(considering that it is the fifth largest rice exporters). Countries like those in western Africa (e.g. Ghana and Ivory Coast) would be highly vulnerable, as they are part of the same community as China (figure 4) and would face the task of competing with China on the global rice markets. International rice prices would increase, assuming that rice production does not increase substantially elsewhere, there is no major release of rice reserves to the international markets (e.g. as Japan did in 2008), and that there major changes to the other global grain markets. In this situation, low- and lower-middle-income countries that are dependent on imports for their staple food supply will be at a severe disadvantage.

The community structure of the soybean network is quite different from the structure of the rice network (figure 5), so we might expect *a priori* that there are differences in shock vulnerability. The soybean network reveals one of the most concentrated community structure, composed by only three large clusters without a clear regional scheme (figure 4). The most important bloc—in terms of trade volume—includes the US and Brazil from the producing and exporting side, which together account for over 70% of global soybean exports, and China from the importing side, which alone accounts for 56% of global soybeans imports. If one of these main producers experiences a sharp decline in production, the global implications of the shock will largely depend on the capacity of few other major producing countries to make up for the production shortfall.

The global wheat market has a community structure that falls in-between the structures found in the rice and soybean markets. Major producers are grouped together in three separate communities: (1) the US, Canada, and Australia, (2) Argentina and Brazil, (3) Russia and Ukraine. Interestingly, Europe belongs to yet another separate cluster, in which France is the notable producer and exporter. One might hypothesize that this geographic diversity is advantageous for dealing with a disruption, particularly if it has as spatial component (e.g. crop disease spreading over an area, a regional conflict, or regional-scale extreme weather). Of course, community structure alone is not sufficient for understanding the impacts of shocks on these global markets.

Knowledge of community structure can be linked to the latest efforts to understand non-equilibrium conditions in the global food system. For example, recent models of food shock propagation [18, 58, 59] would benefit from these community-structure insights. Improved disruption scenarios can be generated to analyze potential responses and identify vulnerabilities of the food system, at scales ranging from the individual country to the global system.

More generally, the role of food price shocks in shaping the community structure of global food-trade system should be better understood [60, 61]. Food

price shocks can alter global trade patterns as they typically encourage countries both to rise export barriers and to lower import tariffs, which may in turn exacerbate price spikes. Such protectionist measures are often combined with other frequent responses such as panic buying, large-scale governmental intervention, hoarding and precautionary purchase. These common short-term remedies associated with price spikes are poorly understood although they may have pervasive consequences on less developed countries, generally extremely dependent on imports, thus altering the way in which they locally form their trade networks.

Along similar lines, one may investigate more deeply the importance of other determinants of bilateral import-export flows in explaining the formation of clusters in the international web of food trade. For example, exchange rate volatility has grown significantly after the GFC. This can correlate with trade growth, as typically the more a country undergoes currency devaluation, the slower the growth in its trade [41]. Other determinants to be explored include climate-related shocks, which are especially relevant because of crop sensitivity to weather extremes [10, 11], regional conflicts, epidemics, agro-terrorism and crop pests [12].

From a more methodological perspective, this study could be improved through additional tests aimed at checking the robustness of the main results against alternative parameterizations of (and assumptions about) the community-detection algorithms employed. For example, the well-known resolution-limit bias affecting many existing methods may be explored using the multiple-resolution community detection strategy by introduced in [62]. Likewise, additional analyses on shortcomings of FAO-STAT bilateral trade data (e.g. possible underreporting of intra-Africa trade) should examine how sensitive community detection and analysis are to systematic trade-data biases. Furthermore, despite the fact that the foregoing analysis was focused on the identification of non-overlapping communities, this work can be extended using community-detection algorithms that look for clusters that may partly overlap [63, 64]. This is important, as knowing the degree of overlap among communities may shed more light on the way in which food crises may spread across clusters. Finally, when analyzing the IFTMN as a multi-layer network, we have implicitly assumed that any pair of layers are linked by fictional edges connecting the same country in the two layers, and that the weights of this edge are homogeneous across countries and equal to one. Such a system parameter, however, may affect the emerging community structure [65]. Therefore, experimenting with different parameter values can give interesting insights into the emergence of clusters in the product-country space.

Acknowledgments

Giorgio Fagiolo gratefully acknowledges support by the European Union's Horizon 2020 research and innovation program under grant agreement No. 649186–ISIGrowth. M J Puma gratefully acknowledges fellowship support from the Columbia University Center for Climate and Life.

ORCID iDs

G Mangioni  <https://orcid.org/0000-0001-6910-0112>

M J Puma  <https://orcid.org/0000-0002-4255-8454>

References

- [1] Porkka M, Kummu M, Siebert S and Varis O 2013 *PLoS ONE* **8** 1–12
- [2] United Nations 2015 Transforming our world: the 2030 agenda for sustainable development *Tech. Rep. A/RES/70/1* (UN General Assembly) (<https://sustainabledevelopment.un.org/resourcelibrary>)
- [3] D'Odorico P, Carr J A, Laio F, Ridolfi L and Vandoni S 2014 *Earth's Future* **2** 458–69
- [4] Godfray H C J, Beddington J R, Crute I R, Haddad L, Lawrence D, Muir J F, Pretty J, Robinson S, Thomas S M and Toulmin C 2010 *Science* **327** 812–8
- [5] United Nations 2015 World Population Prospects: The 2015 Revision, Key Findings and Advance Tables *Tech. Rep. ESA/P/WP.241* (New York: Department of Economic and Social Affairs, Population Division)
- [6] Hazell P and Wood S 2008 *Phil. Trans. R. Soc. Lond. B. Biol. Sci.* **363** 495–515
- [7] Hanjra M A and Qureshi M E 2010 *Food Policy* **35** 365–77
- [8] Woods J, Williams A, Hughes J K, Black M and Murphy R 2010 *Phil. Trans. R. Soc. B. Biol. Sci.* **365** 2991–3006
- [9] Coumou D and Rahmstorf S 2012 *Nat. Clim. Change* **2** 491–6
- [10] Battisti D S and Naylor R L 2009 *Science* **323** 240–4
- [11] Gornall J, Betts R, Burke E, Clark R, Camp J, Willett K and Wiltshire A 2010 *Phil. T. Roy. Soc. B: Biol. Sci.* **365** 2973–89
- [12] McCloskey B, Dar O, Zumla A and Heymann D L 2014 *Lancet Infect. Dis.* **14** 1001–10
- [13] Nonhebel S and Kastner T 2011 *Livest. Sci.* **139** 3–10
- [14] Tilman D, Balzer C, Hill J and Befort B L 2011 *Proc. Natl Acad. Sci.* **108** 20260–4
- [15] Cassidy E S, West P C, Gerber J S and Foley J A 2013 *Environ. Res. Lett.* **8** 034015
- [16] Clapp J and Cohen M J 2009 *The global food crisis: Governance challenges and opportunities* (Waterloo, ON: Wilfrid Laurier University Press)
- [17] Clapp J 2015 Food security and trade: Unpacking disputed narratives *Tech. Report* (Rome: Food and Agriculture Organization of the United Nations)
- [18] Puma M J, Bose S, Chon S Y and Cook B I 2015 *Environ. Res. Lett.* **10** 024007
- [19] Fagiolo G, Schiavo S and Reyes J 2009 *Phys. Rev. E* **79** 036115
- [20] Lee K-M, Yang J S, Kim G, Lee J, Goh K I and Kim I M 2011 *PLoS ONE* **6** e18443
- [21] d'Amour C B, Wenz L, Kalkuhl M, Steckel J C and Creutzig F 2016 *Environ. Res. Lett.* **11** 035007
- [22] Haldane A G and May R M 2011 *Nature* **469** 351–5
- [23] Fagiolo G 2017 *The International Trade Network: Empirics and Modeling* ed J N Victor, A H Montgomery and M Lubell (Oxford: Oxford University Press)
- [24] Barigozzi M, Fagiolo G and Garlaschelli D 2010 *Phys. Rev. E* **81** 046104

- [25] Barigozzi M, Fagiolo G and Mangioni G 2011 *Physica A: Stat. Mech. Appl.* **390** 2051–66
- [26] Ercsey-Ravasz M, Toroczkai Z, Lakner Z and Baranyi J 2012 *PLoS ONE* **7** 1–7
- [27] Shuttters S T and Muneeppeerakul R 2012 *PLoS ONE* **7** e39756
- [28] Brooks D H, Ferrarini B and Go E C 2013 *J. Int. Comm. Eco. Pol.* **4** 1350015
- [29] Wu F and Guclu H 2013 *Risk Anal.* **33** 2168–78
- [30] Lin X, Dang Q and Konar M 2014 *Environ. Sci. Tech.* **48** 5439–47
- [31] Gephart J A and Pace M L 2015 *Environ. Res. Lett.* **10** 125014
- [32] Fracasso A, Sartori M and Schiavo S 2016 *Sci. Environ.* **543B** 1054–62
- [33] Battiston F, Nicosia V and Latora V 2014 *Phys. Rev. E* **89** 032804
- [34] Kivelä M, Arenas A, Barthelemy M, Gleeson J P, Moreno Y and Porter M A 2014 *J. Comp. Net.* **2** 203
- [35] Boccaletti S, Bianconi G, Criado R, del Genio C I, Gomez-Gardenes J, Romance M, Sendina-Nadal I, Wang Z and Zanin M 2014 *Phys. Rep.* **544** 1–122
- [36] Fortunato S 2010 *Phys. Rep.* **486** 75–174
- [37] Winders B 2017 *Grains* (Cambridge: Polity Press)
- [38] Malliaros F D and Vazirgiannis M 2013 *Phys. Rep.* **533** 95–142
- [39] Kim J and Lee J G 2015 *SIGMOD Rec.* **44** 37–48
- [40] Porter M A, Onnela J P and Mucha P J 2016 arXiv:0902.3788
- [41] Kang J W 2016 International trade and exchange rate Working Paper 498 (Asian Development Bank) (www.adb.org/sites/default/files/publication/202841/ewp-498.pdf)
- [42] Billen G, Lassaletta L and Garnier J 2014 *Glob. Food Sec.* **3** 209–19
- [43] Konar M, Dalin C, Suweis S, Hanasaki N, Rinaldo A and Rodriguez-Iturbe I 2011 *Water Resour. Res.* **47**
- [44] Sartori M and Schiavo S 2014 Virtual water trade and country vulnerability: A network perspective *Tech. Report 77* (Trento: Bocconi University)
- [45] Tamea S, Carr J A, Laio F and Ridolfi L 2014 *Water Resour. Res.* **50** 17–28
- [46] Dalin C, Wada Y, Kastner T and Puma M J 2017 *Nature* **543** 700
- [47] Fagiolo G 2006 *Econ. Bull.* **3** 1–12
- [48] Freeman L C 1978 *Soc. Net.* **1** 215–39
- [49] Fagiolo G 2007 *Phys. Rev. E* **76** 026107
- [50] Newman M E J and Girvan M *Phys. Rev. E* **69** 026113
- [51] Arenas A, Duch J, Fernandez A and Gómez S 2007 *CoRR* 0702015
- [52] Barrat A, Barthélemy M, Pastor-Satorras R and Vespignani A 2004 *Proc. Natl Acad. Sci. USA* **101** 3747–52
- [53] Rotta R and Noack A 2011 *J. Exp. Algorithmics* **16** 2.3:2.1–2.27
- [54] Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 *J. Stat. Mech: Theory Exp.* **P10008**
- [55] Winkelmann R 2008 *Econometric Analysis of Count Data* (New York: Springer)
- [56] Baldwin R and Taglioni D 2006 Gravity for dummies and dummies for gravity equations *Working Paper* 12516 (Cambridge, MA: National Bureau of Economic Research)
- [57] Anderson J E 2011 *Annu. Rev. Econ.* **3** 133–60
- [58] Gephart J A, Rovenskaya E, Dieckmann U, Pace M L and Brännström Å 2016 *Environ. Res. Lett.* **11** 035008
- [59] Marchand P *et al* 2016 *Environ. Res. Lett.* **11** 095009
- [60] Headey D 2011 *Food Policy* **36** 136–46
- [61] Anderson K and Nelgen S 2012 *Oxford Rev. Econ. Pol.* **28** 235
- [62] Arenas A, Fernandez A and Gomez S 2008 *New J. Phys.* **10** 053039
- [63] Nicosia V, Mangioni G, Carchiolo V and Malgeri M 2009 *J. Stat. Mech: Theory Exp.* **P03024**
- [64] Xie J, Kelley S and Szymanski B K 2013 *ACM Comput. Surv.* **45** 43:1–35
- [65] Mucha P J, Richardson T, Macon K, Porter M A and Onnela J P 2010 *Science* **328** 876–8
- [66] Carchiolo V, Longheu A, Malgeri M and Mangioni G 2011 *Complex Networks: Second International Workshop, CompleNet 2010, Rio de Janeiro, Brazil, October 13–15, 2010, Revised Selected Papers* ed L da F Costa, A Evsukoff, G Mangioni and R Menezes (Berlin: Springer) pp 187–95
- [67] De Domenico M, Porter M A and Arenas A 2015 *J. Complex Networks* **3** 159
- [68] Danon L, Diaz-Guilera A, Duch J and Arenas A 2005 *J. Stat. Mech: Theory Exp.* **P09008**
- [69] de Sousa J and Lochar J 2011 *Scand. J. Econ.* **113** 553–78