Int J Life Cycle Assess (2017) 22:1146–1158
DOI 10.1007/s11367-016-1222-6

WOOD AND OTHER RENEWABLE RESOURCES

Identifying marginal supplying countries of wood products via trade network analysis

Massimo Pizzol[1] · **Marco Scotti**[2]

Received: 12 May 2016 / Accepted: 3 November 2016 / Published online: 10 November 2016

Abstract
Purpose The consequential inventory modeling approach for life cycle assessment implies that an increase in the demand for a specific product is met by the marginal suppliers within the market. The identification of marginal suppliers is however complicated by difficulties in defining appropriate geographical market delimitations. In this study, an advanced system thinking approach is proposed to address this challenge in the identification of marginal supplying countries of wood products.
Methods Groups of countries which represent geographical markets are identified from trade data by using a network analysis-based clustering technique. Within these markets, marginal supplying countries are selected based on positive historical increments. The analysis covers 12 different products and all countries in the world using trade data for the period 1998–2013.
Results and discussion Global indices allow differentiating how product-specific trade networks are separated into communities and how interconnected these networks are. Large differences between products and minor differences between trade years are observed. Communities identified for each product tend to overlap with existing geographical regions and seem thus realistic. By combining this information with product-specific production increment rankings, marginal supplying countries of wood products were identified.
Conclusions The identified geographical market delimitation is a key for proper consequential life cycle assessment (LCA) inventory modeling in areas such as timber-based construction and biomass-based energy production. The method can in principle be applied to any product for which trade network data are available and ideally should be accompanied by a detailed analysis of technological constrains within the identified supplying country.

Keywords Consequential life cycle assessment · Network analysis · Network clustering · Network communities · Trade · Wood products

1 Introduction

The consequential inventory modeling approach for life cycle assessment (LCA) implies that an increase in the demand for a specific product is met by the marginal suppliers within the market (Ekvall and Weidema 2004). In a stable or increasing market, the marginal suppliers are able to increase the production capacity at the lowest cost. The identification of marginal suppliers is a key element of consequential LCA and defines which activities and relative inputs and outputs should be included in the life cycle inventory, thus determining the final environmental burden of the product system. Previous studies show that LCA results can change dramatically depending on the choice of the marginal supplier. Examples range from electricity supply (Lund et al. 2010; Mathiesen et al. 2009) to production of vegetable oils (Schmidt 2010, 2015).

Responsible editor: Yi Yang

Electronic supplementary material The online version of this article (doi:10.1007/s11367-016-1222-6) contains supplementary material, which is available to authorized users.

✉ Massimo Pizzol
massimo@plan.aau.dk

1 Department of Development and Planning, Aalborg University, Rendsburggade 14, 9000 Aalborg, Denmark

2 GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany

The methodology for identification of marginal suppliers is described in previous studies (Weidema et al. 2009, 1999). The main idea behind the method is that constrained suppliers should be identified and removed from the list of potential suppliers, as they cannot respond to a change in demand. Geography is a constraint because when a market is regionalized, the suppliers far from the point of demand cannot respond to changes. Co-production is another typical constraint because multi-functional processes only respond to changes in the demand of the main product and an increase in the demand of co-product would not result in production increase. Technology can also be a constraint if not accessible or not further exploitable (e.g., hydropower) and thus not affected by demand. Last, policy-related constraints exist such as production quotas or the political decision of phasing out a technology.

Although consequential LCA has been developing substantially in recent years, methodological aspects of marginal suppliers' identification have not been the subject of thorough research. Even if a sound and general methodology exists, it is unclear how consistently applicable it is to different cases. Also, the method is not strict in defining which type of data and sources are the most appropriate. This allows for flexibility but makes the analysis highly case-specific and, if not leading to inconsistencies, makes it difficult to compare across studies. For most products, the main problem is the lack of data regarding the geographical delimitation of markets, the future production outlooks or historical trends, and the trade. Consequential LCA has been criticized for the lack of testing of the robustness of its results and has been compared to a scenario analysis exercise (Dale and Kim 2014; Zamagni et al. 2012).

The objective of this work is to develop a network analysis-based approach for geographical market delimitation that can be applied consistently to different products and based on data of similar type. The idea suggested in this work is combining a trade network analysis that identifies clusters of countries historically linked in more intense exchanges of products with a regression analysis of historical production trends for product types across countries. This is proposed as an innovative way to improve marginal supplier identification regarding the geographical constraints. Such an approach might prepare to and ease the successive analysis of technological constraints. The work takes the case of wood products as point of departure.

Various LCA studies have been published on wood products, some of which use the attributional modeling approach (Eshun et al. 2010; Neupane et al. 2011; Schaubroeck et al. 2013) and some others the consequential approach, e.g., for identifying marginal suppliers of wood pulp (Reinhard et al. 2010) and in the hybrid LCA of wood biomass for energy production (Grinde 2011). Wood products are used in several different sectors, e.g., timber as construction material, pulpwood for paper production, and chips and pellets for energy supply. Modeling life cycle inventories for wood products requires knowledge of the type of forest management activities under analysis, e.g., plant species and their growth rate, rotation time, and thinning practices, as these aspects represent the key to account for the amount and timing of carbon emissions and sink (Helin et al. 2012; Jørgensen and Hauschild 2013; Levasseur et al. 2010). The timescale of wood production activities varies substantially depending on their intended product, e.g., spruce to be used as timber construction material may take 70 years to grow, whereas the production of eucalyptus for energy purposes follows shorter rotation times (e.g., 6 years; see De Rosa et al. 2016). Wood co-products are generated at all processing stages: from thinning produced in the forest, to chips and cutovers of sawmilling and residues of the wood pulp production process. Many of these are down-cycled to products of lower economic value, e.g., the less valuable planks from the timber production are used to build fences and pallets. Therefore, markets for woody raw materials, wood products, and wood residues exist. The economic value of wood products and the price can range substantially (e.g., timber versus chips), and such products are transported and traded worldwide. In order to develop an accurate life cycle inventory modeling, it is therefore important to identify where the marginal suppliers of wood products that can respond to the increase in demand are located. The hypothesis used in this study is that network analysis of wood product trade can help in this sense.

Network analysis is nowadays applied in several research fields ranging from social sciences (Wasserman and Faust 2016) to ecology (Fath et al. 2007), resources management (Bodini et al. 2012; Pizzol et al. 2013), and molecular biology (Nguyen et al. 2011; Scott-Boyer et al. 2016). It is particularly interesting in the context of this article to consider previous studies applying network analysis on trade data, i.e., dealing with the study of trade networks. Trade networks are constructed from data of import and export between countries, these quantities being measured either in monetary or physical units. Previous studies have shown that network analysis can be applied successfully to identify system-level trends in, e.g., the globalization of trade (De Benedictis and Tajoli 2011) and in growth and development of complex economic systems (Hausmann and Hidalgo 2011; Hidalgo and Hausmann 2009; Huang and Ulanowicz 2014). Also, network analysis studies in relation to LCA have been published in recent years to investigate structure and complexity of life cycle databases (Heijungs 2012; Navarrete-Gutiérrez et al. 2015). Network analysis has also been applied in the field of LCA to study the complex supply chain of products, e.g., electricity (Kim and Holme 2015) and fuels (Singh and Bakshi 2011), and forests (Schaubroeck et al. 2012, 2013). A thorough review of network analysis applications in the field of input-output analysis is provided by Nuss et al. (2016).

Detailed objectives of this study are therefore (i) to develop a network analysis-based method to identify and delimit geographical markets based on trade between countries, (ii) to test the method on the case of wood products and discuss its advantages and limitations, and (iii) to provide resources for the identification of marginal supplying countries of wood products all over the world, thus allowing to build accurate life cycle inventories for use in consequential LCA.

2 Materials and methods

Yearly trade and production data on all wood products, for all countries in the world, were retrieved from the statistics database of the Food and Agriculture Organization of the United Nations (FAOSTAT 2016). The trade data were used in a network analysis where the *spin glass* clustering algorithm (Reichardt and Bornholdt 2006) was applied to identify communities (i.e., topological clusters) representing geographical markets. The production data were used to calculate historical increments in production for each wood product in each country. Finally, the two types of data were combined to identify the marginal suppliers in each market.

FAOSTAT data are aggregated at country level, and this is the reason why the study identifies marginal supplying countries. This, however, allows to subsequently model in accurate way the wood production life cycle inventories, once the marginal supplying countries have been identified. Moreover, country-specific technological and political constraints may affect the trade, and their effects are considered to be already embedded in the final trade and production data reported at country level. Since LCA background inventory modeling is currently done at country level, this scale was considered appropriate for the intended use of the results.

2.1 Cleaning and preparation of network data

Since transport costs and other geopolitical factors may limit the trade of different wood products to specific areas, then geographically separate markets (i.e., groups of countries representing different trade "communities") can in principle be identified using clustering algorithms from network analysis. Markets are defined as "geographical" because this study focuses on how geography affects trade as the main factor for market delimitation (other factors such as technology, co-production, and policy are not directly taken into account) and investigates which countries belong to the same market and thus what the market delimitation is. Countries affiliated to a community are more closely related to the countries within the community than to others not comprised in the community. In this context, being related means "trading," both in terms of number of countries between which mutual trade occurs and amount of product traded.

A network is constituted of *vertices* connected by *edges*, the value of each edge being its *weight* and where connections may be *directed* if the exchange of any *currency* between vertices follows a specific direction (Borgatti et al. 2009). The trade data retrieved from FAOSTAT report the amount of wood products that is exchanged (imported to and exported from) between countries, and this can be defined as a weighted, directed network (each country represents a vertex of the network, and exports consist of directed edges between pairs of vertices, i.e., from exporter to importer). A different network can be identified for each product and for each year. These networks are mono-modal, and all flows are measured in mass units, where the specific unit depends on the product under analysis. In other words, each network is based on a single currency. Since the data are artificially balanced and the trade balance differs slightly when looking at imports and exports, data cleaning was required: (1) Export quantities were chosen as edge weights, and these trade flows were organized into a material flow matrix. If a country exports to another and vice versa, both export quantities are included in the analysis as two different network edges. Export quantities are intended by FAOSTAT as "*Products of domestic origin or manufacture shipped out of the country*"; they include re-exports but exclude "in-transit" shipments (FAOSTAT 2016). (2) Vertices not representing countries (e.g., "world" and "balance") were removed. (3) Edges with weight lower than a certain cutoff value were removed. (4) Isolated vertices were removed, i.e., vertices for which the number of connections (*degree*) was lower than one. (5) Isolated pairs of vertices or small subnetworks were removed as the identification of clusters was carried out using the giant component only, i.e., the largest connected component of the whole network. Preliminary analysis showed that isolated vertices and pairs of vertices usually represent small countries with limited trade in wood products and are in fact isolated because of the cutoff applied. Thus, the choice of excluding them was not considered to affect results substantially. What is described in the last three steps is in agreement with the approach that Nguyen et al. (2011) have previously called "focusing on the giant components." The data cleaning and further cluster analysis were performed in the R Statistical Environment (R Core Team 2005).

Without cutoff, all trade networks are almost entirely interconnected, because almost all countries trade between each other (even if, often, small quantities only). Since the focus is on the large quantities that countries trade with their close commercial partners, i.e., within their community, cutting off the network reduces the bias towards weak links that could impair the sound identification of network communities. A network-specific cutoff value was calculated for each network, because networks are in different units and the range of edge weights differs between networks. The cutoff value was calculated as the geometric mean of the weights in the complete network, based on the observation that weights are approximately log-normally distributed. Cutting off the network reduces its total weight size minimally, as more than 99% of the total weight is usually maintained (see Electronic Supplementary Material). This procedure removes part of the edges (i.e., the ones responsible for exchanges of smaller magnitude), thus increasing the fragmentation of the network and improving the identification of communities.

The network data cleaning was performed for each product and each year in the period 1998–2013. Thus, 384 networks were constructed: 12 products × 16 trade years × with/without cutoff. The products under analysis are identified by the following names: *chips and particles*, *fiberboard*, *Ind Rwd Wir (C)*, *Ind Rwd Wir (NC) other*, *Ind Rwd Wir (NC) Tropica*, *newsprint*, *particle board*, *plywood, Sawnwood (C)*, *Sawnwood (NC)*, *veneer sheets*, and *wood pulp*, where Ind = industrial, Rwd = round wood, Tropica = tropical, Wir = wood in the rough, C = coniferous, and NC = non-coniferous. The unit is cubic meters solid volume (m^3) for all products except wood pulp, which is measured in metric tons air dry weight (i.e., with 10% moisture content), and newsprint, which is measured in metric tons. A detailed explanation of what each product name refers to and of the units can be accessed from FAOSTAT (2016).

2.2 Detection of communities representing geographically separate markets

The *spin glass* method proposed by Reichardt and Bornholdt (Reichardt and Bornholdt 2006) was applied to partition the network into communities. In particular, the version of this method that is implemented in the function "*cluster_spinglass*" of the R library *igraph* (Csardi and Nepusz 2006) was used. This method represents a widely applied algorithm for community detection, i.e., a technique to identify groups of densely interconnected vertices in a network (Newman and Girvan 2004), and it is widely used. The algorithm is based on the idea that vertices in the same network community exchange most of their currency within the community than outside it. The spin glass method was applied in its weighted version: Community affiliation does not merely depend on the number of exchanges but also considers the amount of currency exchanged. The choice of using the spin glass method is motivated by its robustness. Indeed, such an approach has been shown to successfully detect network communities in very different fields ranging from molecular biology (Caberlotto et al. 2013) to sociology (Rodriguez and Pepe 2008) and economics (Reichardt and Bornholdt 2007).

The spin glass method was applied on cleaned network data and used to detect the communities. The procedure identifies different network communities each time the function is iterated. In order to deal with the variability of results, a probabilistic approach was adopted, inspired by a previous study dealing with the same problem (Caberlotto et al. 2013). First, the clustering function was iterated 100 times on each of the 384 networks, and 38,400 different sets of communities were identified. Then, a further aggregation of all clustering results was performed, and product-specific contingency tables were derived. In the contingency tables, the frequency of appearance of each possible pair of countries in the same community is summarized.

Two system-level indicators were calculated for each network, the most important being *modularity* (Q). Modularity is a global property that can be used to quantify the goodness of a specific division of the network into communities (i.e., densely connected subgroups of vertices; see Fortunato 2010). The most popular approach for calculating Q is the one proposed by Newman and Girvan (2004): Q is determined by comparing a network with its randomized versions, i.e., networks that have the same number of edges and vertices but where the edges between the vertices (and eventually their weights) are simulated at random. Q serves as a quantitative criterion for evaluating whether the modular structure of the original network significantly deviates from the average of its randomized versions. When the fraction of within-community connections is no different from what expected with randomized networks, then Q is zero. Non-zero values represent deviations from randomness. Q values above 0.3 are considered to indicate significant community structure in the original network (Clauset et al. 2004). In general, increasing values of Q indicate better partition of the network into clusters (Newman and Girvan 2004). In this study, both a weighted and a non-weighted version of Q were calculated. The former takes into account the strength of the connections between vertices, i.e., the amount of product traded between two countries. The latter only considers the presence (or absence) of edges in the network. This index does not take into account the directionality of edges. The second system-level indicator calculated was *density* (γ). Density is computed as the ratio between existing edges and all possible interactions, and it is used to quantify how interconnected the network is (De Benedictis and Tajoli 2011). If all vertices are connected, then the network is said to be complete and $\gamma = 1$. Differently from Q, γ is calculated disregarding the network weights and thus describes the global structure of the network only.

In summary, 38,400 Q values (both weighted and non-weighted) and 384 γ values were calculated. Statistical tests were then conducted to study the differences between modularity values calculated with or without cutoff for different years and taking into account the variability across products. Testing for the effect of the cutoff is useful to investigate the type of bias introduced by data cleaning, while if modularity changes significantly as a function of the years, this may reflect variations in the trade network structure. Randomized complete block design (RCBD) ANOVA was firstly applied to compare the effect of cutoff and year (and their interaction) on modularity, using products as blocking factor. The ANOVA was performed both on the mode and the mean Q values calculated on the 100 iterations and testing either weighted or non-weighted Q values. Pairwise comparisons with the Tukey's honest significant difference (HSD) test were performed to identify the specific differences between years and products.

2.3 Selecting marginal supplying countries in each market based on historical production increments

In a growing market, the suppliers that can respond to a change in demand are those that can expand their production capacity. One way of expressing this quantitatively is by calculating the expected absolute production increment over a certain period. Future outlooks can be used as source for this information but are intrinsically highly uncertain. In this study, historical data were therefore used. For each wood product and each country in the database, a linear regression analysis was conducted using time series of production data. The slope of each linear regression represents the yearly production increment in the period. For most products, the production time series for the period 1998–2013 were available with minimal data gaps and were therefore used to calculate historical increments. Countries were then ranked from the largest positive to the largest negative increment. When an increase in demand occurs, then it is assumed that the trade between countries will balance it out (i.e., cascade effects) and the suppliers located in the country with the highest capacity will meet the demand. The group of suppliers with the highest capacity here belongs to the country with the largest yearly production increment (averaged over the period 1998–2013). This means that a systemic perspective is applied. Since countries are connected in the trade network, when a change in the demand occurs in one geographical market, then the suppliers of all countries in such market (i.e., in the same network community) will be affected and in particular those in the marginal supplying country.

3 Results

3.1 Communities representing geographically separate markets

The network analysis identified communities representing geographically separate markets for each wood product and each year. In some cases, a certain match was observed between the network communities and the spatial distribution of countries, so that communities overlap with specific trade areas (e.g., the European Union) or continents, and this was taken as a proof that the clustering algorithm returns realistic results. In other cases, the grouping proposed by network clustering deviates from a simple spatial analysis, and this justifies the use of network analysis as a tool to identify separate markets. Results are case-specific and differ depending on the product and the year under analysis. Figure 1 shows an example of communities detected for the product Sawnwood (C) in 2013. The full data on cutoff values, percent of network weight maintained after cutoff, modularity, and density for all networks (all products and all years) are provided as Electronic Supplementary Material.

The ANOVA showed minor or no significant differences between years and significant differences between products in all cases (Table 1). The Tukey HSD test showed that the minor significant differences between years are only due to 2007, being this year different from the others for weighted Q. The use of the cutoff affected in a significant way the analysis and biased both modularity scores. This means that imposing the cutoff, i.e., removing the weakest edges, influences the computation of the weighted modularity, even if the major trade flows are preserved. As expected, using the cutoff increases the calculated modularity values. However, no significant interaction was observed between cutoff and trade year, thus indicating that the choice of using a cutoff did not create or hide differences between years. The networks obtained with the cutoff represent the reference for further interpretation of results. When performing the ANOVA on the reduced dataset (i.e., taking into account the Q values of networks obtained with the cutoff only), the results are consistent with those described above. The larger differences between products compared to years can be intuitively appreciated looking at Fig. 2. This figure shows the weighted Q values calculated with the cutoff across products and years. Detailed ANOVA summaries, means of all groups, and Tukey HSD results are reported in the Electronic Supplementary Material.

The ranking of products in terms of mean, weighted Q (in parenthesis) calculated over years and iterations is particle board (0.514), Ind Rwd Wir (NC) other (0.513), newsprint (0.436), Ind Rwd Wir (C) (0.426), fiberboard (0.42), Sawnwood (C) (0.408), veneer sheets (0.397), chips and particles (0.378), Sawnwood (NC) (0.337), plywood (0.288), wood pulp (0.243), and Ind Rwd Wir (NC) Tropica (0.230). Yearly values are shown in Fig. 2. Modularity values higher than 0.3 indicate a significant fragmentation of the network. The higher the value of Q, the more clear-cut the separation of countries into submarkets (i.e., the structure of the world market is more fragmented). This does not necessarily mean that many markets are identified but simply stands for a network structure with overt clusters (i.e., the global market may just be split into two markets, and Q only tells us how strongly these markets are separated). High Q values indicate strong separation; low indicates weak separation. In the case of strong separation, there is high intra-community trade and low inter-community trade, i.e., countries within one community have limited trade with the countries of another community. In the case of weak separation, the inter-community trade

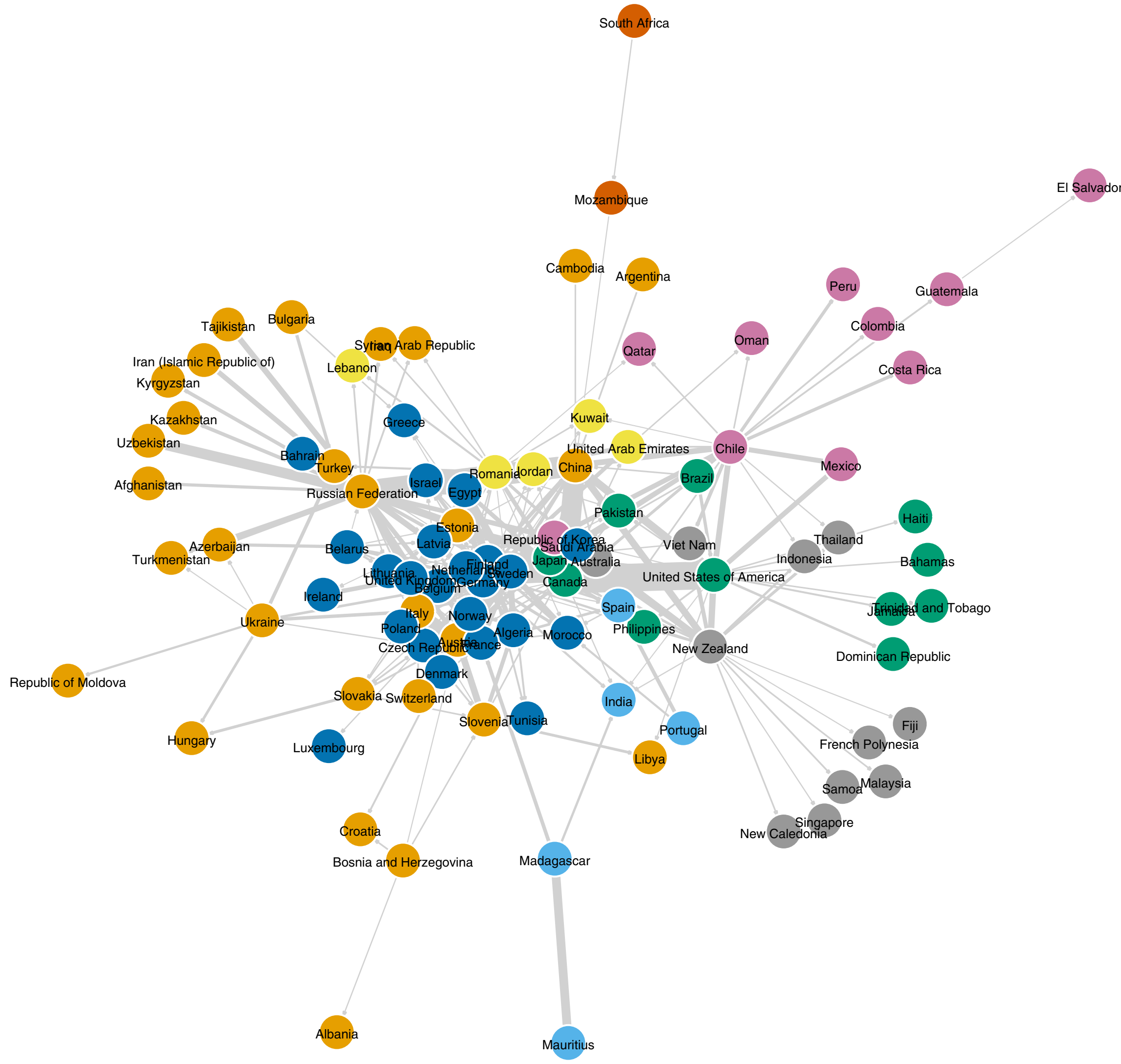


Fig. 1 Communities identified for the trade network of *Sawnwood (C)* in 2013. Edge weight is proportional to the amount of products exchanged among countries, and the direction of exchanges is indicated by *arrowhead edges*. *Colors of the vertices* reflect the affiliation of each country to network communities

is substantial. This latter case could be assumed to represent a globalized trade.

Figure 3 shows product-specific yearly density values. The order of products in terms of mean density (in parenthesis) calculated over years is wood pulp (0.087), Sawnwood (NC) (0.066), veneer sheets (0.065), fiberboard (0.064), particle board (0.055), Sawnwood (C) (0.050), plywood (0.050), chips and particles (0.048), newsprint (0.041), Ind Rwd Wir (NC) other (0.040), Ind Rwd Wir (C) (0.039), Ind Rwd Wir (NC) Tropica (0.035). The trade networks of wood pulp and Ind Rwd Wir (NC) Tropica have similar modularity (lower if compared to the other products), but the former has higher density and is more interconnected. The trade network of wood pulp consists of a global network with many producers and exports, while the trade network of Ind Rwd Wir (NC) Tropica is a global network with few producers. Similarly, the trade networks of Ind Rwd Wir (NC) other and particle board have similar modularity values that are higher if compared to the other networks. The trade network of Ind Rwd Wir (NC) other has lower density than the one of particle board, thus being less interconnected: It represents a network with separate markets centered on a few traders, while the trade network of particle board shows separate markers with many traders.

Product-specific contingency tables reporting the frequency of appearance of two countries in the same community are provided as Electronic Supplementary Material. These square tables are symmetric. They can be used to select a country of interest (e.g., where the demand for wood products is occurring) and determine the likelihood of finding other countries in the same market, in order to identify the marginal supplier among these countries. Descriptive statistics for the contingency tables are provided in Table 2. Figure 4 illustrates the

Table 1 Results from the application of repeated completed block design (RCBD) analysis of variance (ANOVA)

RCBD ANOVA	Weighted modularity	Non-weighted modularity
Mode values of 100 iterations	Cutoff	Cutoff
	$F(1, 341) = 31.17, p < 0.001$	$F(1, 341) = 766.8, p < 0.001$
	Year	Year
	$F(15, 341) = 1.575, p > 0.050$	$F(15, 341) = 1.659, p > 0.050$
	Cutoff × year	Cutoff × year
	$F(15, 341) = 0.291, p > 0.500$	$F(15, 341) = 0.724, p > 0.500$
	Product	Product
	$F(11, 341) = 76.53, p < 0.001$	$F(11, 341) = 88.51, p < 0.001$
Mean values of 100 iterations	Cutoff	Cutoff
	$F(1, 341) = 37.33, p < 0.001$	$F(1, 341) = 1180, p < 0.001$
	Year	Year
	$F(15, 341) = 1.568, p > 0.050$	$F(15, 341) = 1.197, p > 0.100$
	Cutoff × year	Cutoff × year
	$F(15, 341) = 0.182, p > 0.500$	$F(15, 341) = 0.330, p > 0.500$
	Product	Product
	$F(11, 341) = 93.76, p < 0.001$	$F(11, 341) = 141.8, p < 0.001$

The table reports the significance of the effect of trade year, type of product, and cutoff on mode and mean modularity calculated over 100 iterations and the interaction effect between cutoff and year on modularity. Either the weighted or the non-weighted version of modularity is considered as dependent variable

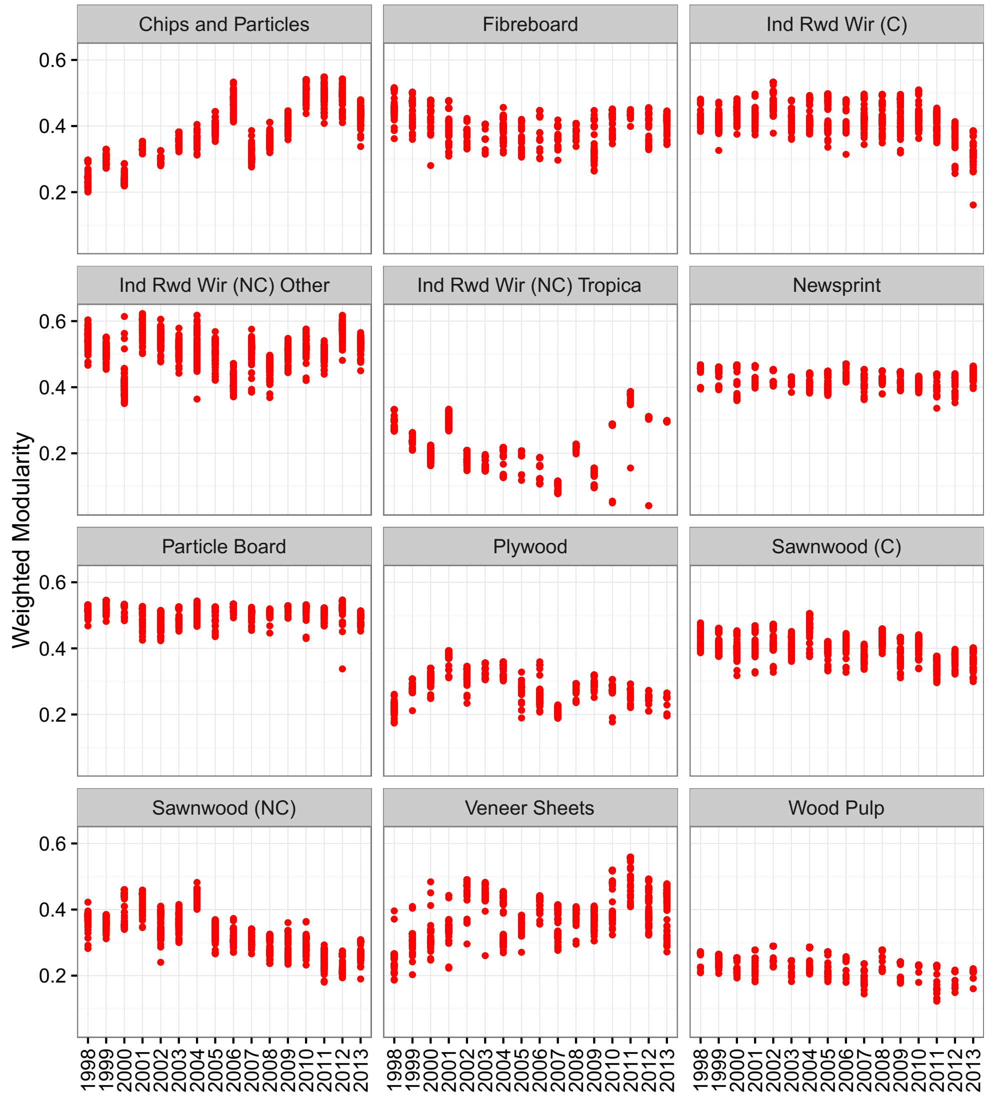
heat map of each contingency table. These results show that in the trade networks of wood pulp, the mean frequency of appearance of a country in a community (318 out of 1600) is almost twice as high as in the case of chips and particles (158 out of 1600). However, the number of countries that were assigned at least once to a community via the clustering method is similarly low between the two products (123 and 129, respectively, out of 171 countries). This indicates that although it is difficult to split the trade network of these products into communities (thus confirming the low Q values of Fig. 2), this may be due to opposite reasons, like, e.g., a high interconnectedness in the case of wood pulp and low interconnectedness in the case of chips and particles. This interpretation is supported by the higher density of the wood pulp trade network compared to the chips and particles network (Fig. 3).

3.2 Historical production increments and marginal supplying countries

The top ten production increment countries for the product Sawnwood (C) are reported in Fig. 5 as example, whereas detailed rankings for each wood product are reported in the Electronic Supplementary Material. Figure 5 shows that large countries such as China and Russian Federation have the largest yearly production increment in absolute value; this means that not only these countries have been increasing the production in the last 16 years but also that the magnitude of this increment is the highest among all countries. The larger the confidence intervals of the slope value, the higher the variability of increments in production across years (these can vary both in absolute value and sign; e.g., see the confidence interval of Sweden that includes negative values).

The identification of marginal supplying countries is performed combining information from the product-specific contingency table with information from the product-specific production increment ranking (both provided in the Electronic Supplementary Material). Starting from the contingency table, the main problem is deciding the minimum acceptable frequency for including a country into the geographical market. The number of countries with high frequency differs according to the product taken into account, so the analysis should be product-specific. A pragmatic solution to this problem was to take a low value as threshold for a country to be included in the geographical market (e.g., 10% of the maximum frequency) and then include in the marginal supplying countries' mix the first three to five countries of the marginal production increment ranking. An example of identification of marginal supplying countries is here provided for the case of Sawnwood (C) demand in Denmark. The row/column named "Denmark" in the contingency table for Sawnwood (C) cointains 167 elements, 68 of which have a value higher than 160 (i.e., higher than 10% of the maximum frequency possible for the row/column, which is 1600). The corresponding countries are assumed to be the geographical market for Sawnwood (C) demand in Denmark. The first five countries listed in the marginal production increment ranking for Sawnwood (C) and included in the geographical market of Denmark are Russian Federation, Germany, Turkey, Sweden, and Belarus. Then, suppliers located in these countries are assumed to be the marginal suppliers of Sawnwood (C) demand in Denmark.

Fig. 2 Values of weighted modularity calculated for the trade networks of different products and trade years (results from all iterations are included). Modularity is calculated on the networks obtained after the application of the cutoff



4 Discussion

In this study, network analysis was applied to trade data of wood products, with the purpose of identifying global and regional markets. To this aim, a network clustering technique was applied using both topology and edge strength for identifying the network clusters. By combining these results with the analysis of production data, it was possible to identify marginal supplying countries of wood products. This information can feed consequential LCA studies on, e.g., woody biomass-based energy production and wood-based construction materials.

A limitation of this model lies in the assumption that communities represent geographical markets and that the clustering algorithm can effectively be used to define how local or global these markets are. This assumption can hardly be validated quantitatively, but a series of qualitative considerations can be made to support it. Theoretically, a market is an arena where commerce is conducted, so a network cluster defined based on trade data between countries should be formally equivalent to the concept of geographical market. The clustering algorithm applied in this study has been validated using various data types, and the clusters correspond to groups of vertices (e.g., enzymes in a molecular network or people in a social network) that interact with each other in a more intimate way than with the rest of the system (e.g., see Caberlotto et al. 2013; Rodriguez and Pepe 2008). Such an approach has not yet been applied to trade data. Previous studies of trade networks have only focused on whole-system indicators (De Benedictis and Tajoli 2011; Huang and Ulanowicz 2014). In network clustering, vertices affiliated to a community are more closely related to vertices within the community than to others outside the community. Two countries may be closely related either because there is a mono-directional exchange of products in substantial amounts or because there is a mutual exchange, or both. This does not exclude that a country

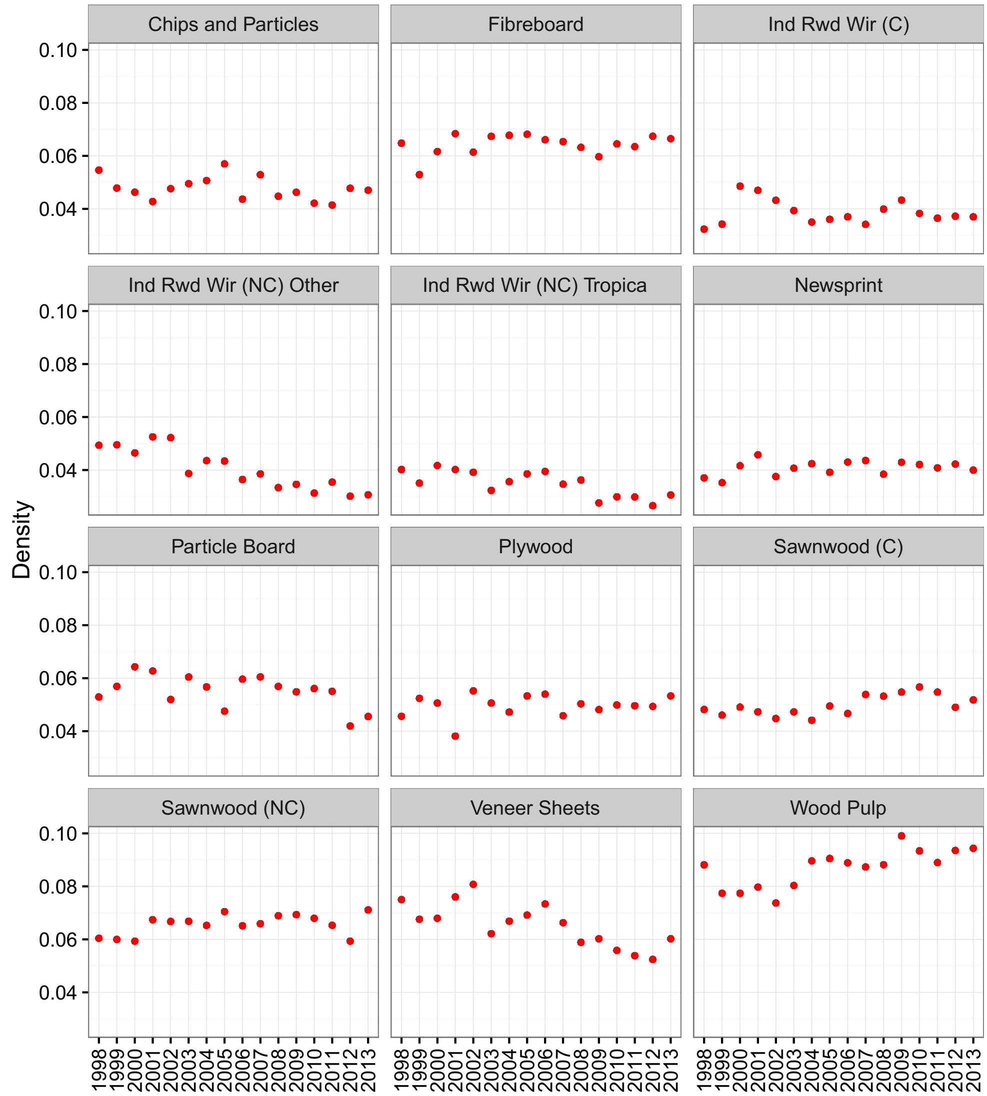


Fig. 3 Values of non-weighted density displayed by the networks that illustrate trade exchanges of different products and trade years. Density is calculated on the networks obtained after the application of the cutoff

belonging to a market (i.e., a community) trades with other countries outside it. The clustering approach to trade is based on network analysis and differs from the approach of economic gravity models (Bergstrand 1985). Indeed, the network-based clustering does not try to predict trade or explain the causes of countries being grouped into communities by using parameters beyond traded quantities (e.g., it does not rely on gross domestic production).

When looking qualitatively at the data, e.g., by analyzing the composition of the communities, the results of the clustering seem reasonable from a geopolitical perspective. It is possible to identify a certain overlap between the markets identified and continents or parts of continents and trade unions (e.g., the Western Europe block and the Asian market). Moreover, when looking at the contingency tables, neighboring countries are generally more likely to appear in the same community. Results showed that the size and the qualitative composition (i.e., affiliated countries) of communities change over time (although the modularity does not change in a significant way for most of the cases: see the results of ANOVA in the Electronic Supplementary Material). Contingency tables were constructed using community affiliations detected through different iterations of the clustering algorithm, and by setting a threshold frequency, it was possible to delimit the various markets. Tests conducted using such threshold-based approach for specific cases showed that results are realistic. However, supporting the market delimitation choice with additional qualitative information (e.g., knowing that a certain tree species is not produced in one country) may prove being a more useful approach than developing a blind quantitative method to set a strict threshold.

Alternatives to historical increments could be used for selection of marginal suppliers. The approach of linear regression on historical production data was chosen since it was simple and provided a unique value (i.e., the slope of the regression line). In this way, suppliers could be easily ranked.

Table 2 Descriptive statistics calculated for the product-specific contingency tables

	NCO	Mean	SD
Chips and particles	123	158	268
Fiberboard	162	292	377
Ind Rwd Wir (C)	168	150	254
Ind Rwd Wir (NC) other	171	192	251
Ind Rwd Wir (NC) Tropica	154	185	236
Newsprint	157	303	408
Particle board	161	215	312
Plywood	177	347	341
Sawnwood (C)	167	201	297
Sawnwood (NC)	173	193	253
Veneer sheets	168	202	267
Wood pulp	129	318	367

The contingency tables report the frequency of appearance of two countries in the same community over all 1600 simulations (16 years × 100 iterations). Mean = average of all frequencies

NCO number of countries (the table includes each country that was assigned at least once to a community), *SD* standard deviation of all frequencies, *Ind* industrial, *Rwd* round wood, *Wir* wood in the rough, *C* coniferous, *NC* non-coniferous

Moreover, the data from historical production were retrieved from the same source of the trade data, so the method was consistent. Even though the use of linear regression analysis on time series is state of the art in LCA studies focusing on marginal supplier identification (Deng and Tian 2015; Schmidt 2015), the approach is rudimentary and here is only proposed as a pragmatic choice for ranking countries within the geographical market. The backbone of this work is the network analysis, and the simplistic choice of a linear regression model was considered acceptable within the context of this study but should undergo further refinements for future applications. A major problem of applying linear regression on the production data is that since results depend on the time period considered for the analysis, the robustness of the results could be weakened by extending the study to a larger amount of historical data or by integrating the dataset with information up to the current production data. Indeed, choosing another time span than 1998–2013 as production data input would change the ranking of countries, although this effect may not be equally noticeable for all wood products under analysis. The period 1998–2013 was chosen because it allows for the largest homogeneous data coverage across countries, as data prior 1998 are unavailable for several countries, and thus was the largest and most consistent dataset available. The uncertainty in the calculated slope varies also substantially across countries as testified by the confidence intervals provided. Performing a sensitivity analysis with respect to the regression-related modeling choices is beyond the scope of this study, but the limitations of the linear model are here acknowledged. A first improvement should be to use more advanced regression techniques for the analysis of time series (Enders 2014). Also, the use of production data as a proxy for production capacity may not be universally accurate, e.g., in the case of countries that are approaching the full capacity following historical increases in production. Qualitative and quantitative information on, e.g., land availability, potential for improvements in production efficiency, and national policies or expected future trade agreements should be considered to complement the analysis and provide a more accurate description of future scenarios. Production costs were not considered in this study, under the simplistic assumption that the suppliers with the largest production increment are also the most economically successful, i.e., the ones with lowest production costs. FAOSTAT provides monetary values of production and trade quantities, but the coverage is not as complete as that of quantities in physical units, and these data were disregarded. However, this information may be available from other sources (e.g., United Nations statistics) or from FAOSTAT in the future and could be used to determine rankings of production increments in the same way as with physical units. The use of historical data for making future predictions is inherently uncertain as there is no guarantee that the future will look like the past. Outlook data may be used instead of historical data even though this may not reduce uncertainty. For example, the forecasts of existing outlook studies on wood products disagree as some see the wood sector at a turning point and consider historical data not useful in building realistic pictures of the future, while others accept the use historical data to forecast future trends (Hetemäki 2014; Hurmekoski 2016; Hänninen et al. 2014).

The information provided in this study can be used to build product-specific and country-specific life cycle inventories similar to those currently provided by commercial LCI databases. Once the geographical market and the marginal supplying countries inside each market have been identified for a specific product, the LCA practitioner can build country-specific inventories for its production. This might mean modeling the country-specific electricity mix and modeling country-specific forest management practices, cf. De Rosa et al. (2016). For example, the major product flow inputs to the wood production process are diesel and electricity. Electricity production is traditionally modeled using a country-specific mix in LCA, and the information about the identified supplying countries could be sufficient for this step. Countries located in different areas have different climate, and this influences the type of wood that can be produced. Even though in principle a main set of species or plant types can be identified for each country, there are notable exceptions of large countries spanning across multiple climate zones (e.g., Russia, China, and USA). Particular attention should be dedicated to assume as homogeneous the management practices adopted in the same country. Thus, the assessment of the

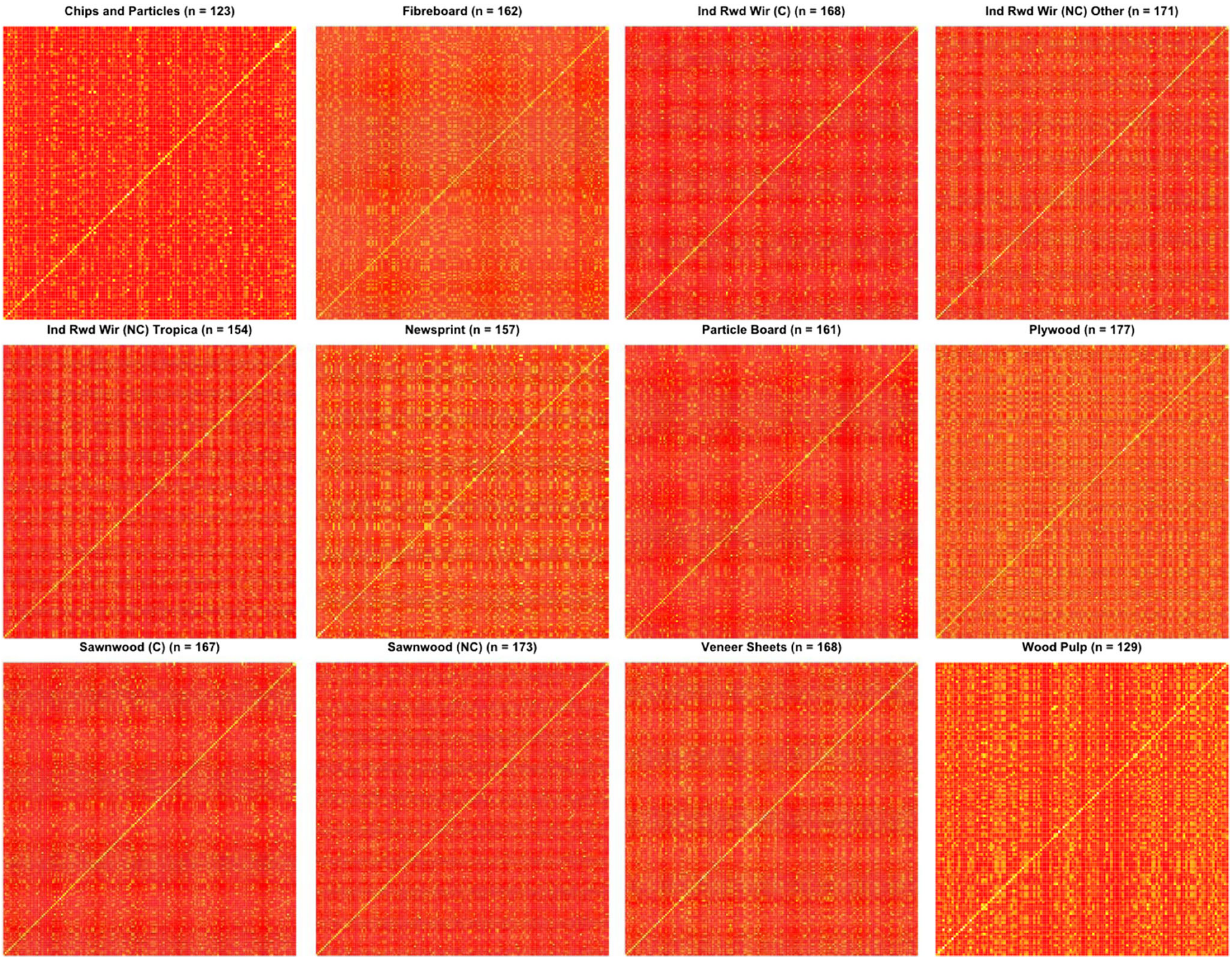


Fig. 4 Product-specific heat maps that visualize the frequencies of occurrence of each country in the network communities. *Yellow color* indicates higher frequency; *red color* indicates low frequency. Max frequency is 1600 (16 years ×100 iterations)

geographical constraints presented in this work should ideally be followed by the assessment of the technology constraints within the marginal supplying countries. The approach here presented can only identify the country where the supplier is located, not the specific technology used. The supplier may use one or a mix of technologies, and this should be duly identified in compiling the inventory. Qualitative and quantitative information on technology-specific production data (e.g., the estimated historical, current, or future share of different technologies on total production in the country) can be used to build accurate supplier-specific inventories. Carrying out this additional step was outside the scope of this study.

The network analysis results proved to be robust despite the variety of wood products (and trade years) under analysis and despite the choices made in data cleaning and preparation (see the ANOVA in Table 1). Beyond the case study of wood products, which was used as an illustrative example, the objective of this article was the development of a consistent method for the identification of marginal supplying countries. Such method can in principle be applied to any product for which trade data are provided in material flow matrix format and at steady state and allows defining markets for the selection of marginal suppliers. Iterating the community detection algorithm several times allows producing contingency tables and increases the robustness of the results. Contingency tables summarize the frequency of occurrence of two countries in the same community and allow studying the variability of results across years. Further testing of the method on products other than wood is required before making a specific evaluation of its general applicability, but some qualitative considerations can already be provided. An interesting question in this respect is whether the method is relevant for cases where technology constraints are expected to affect the environmental burden of a product more than the geographical constraints (i.e., when it is not important to know where the product comes from but what technology the product is made with). Since different countries possess different technologies and industrial infrastructures, the technology and geography

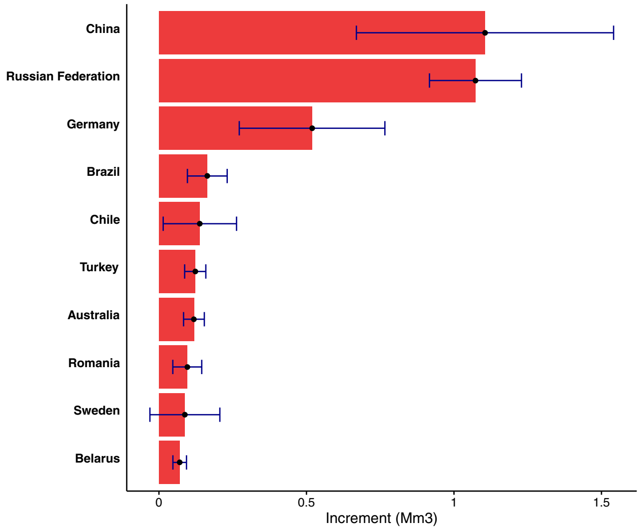


Fig. 5 Top ten countries ranked according to their historical marginal production increment in coniferous *Sawnwood (C)* production, in million cubic meters (Mm^3), calculated for the period 1998–2013. *Red bars* represent the slope of the linear regression calculated on the historical production data; *error bars* represent 95% confidence intervals on this value

constraints are related. So, if the marginal suppliers of, e.g., steel are located in China rather than in Italy, even assuming the technology used by Chinese and Italian producers is identical (e.g., basic oxygen furnace); the environmental burden of the Chinese and Italian steel will be different due to the different electricity mix used in the two countries. Thus, identifying where the marginal suppliers are located via the method would be an important step towards accurate life cycle inventory modeling of the product's production process. A direct application of this method could be used in input-output life cycle assessment. Recently developed multi-regional input-output databases such as Exiobase (Tukker et al. 2013; Wood et al. 2015) allow extracting material flow matrices for a variety of products. Similar data can also be found in the BACI world trade database (CEPII 2016). A recent example where network analysis is applied on monetary input-output tables is the structural analysis of aluminum performed by Nuss et al. (2016).

5 Conclusions

This work has taken wood products as reference case study to show how a system thinking approach can be applied to trade networks for identifying geographical markets and marginal suppliers for consequential LCA. The study is based on the application of a clustering algorithm to trade network data. This application represents a step to improve the current practice for the identification of marginal suppliers in the context of consequential LCA. Even though the proposed approach presents rooms for improvements, it might stimulate the discussion on the topic of marginal supplier identification in the scientific LCA community. Further research should focus on refining and improving the robustness of the regression-based method for the identification of marginal production increments on time series data. Furthermore, the application and testing of this method to the study of other products is recommended, since trade data are available already.

Acknowledgements Thanks to Matthias Buyle, Bo Weidema, and Stefano Merciai for providing insightful and useful comments on draft versions of this manuscript. We acknowledge three anonymous reviewers for the constructive suggestions and the stimulating discussion. This work was funded by the research grant no. 1305-00030B of the Danish Strategic Research Council.

References

Bergstrand JH (1985) The gravity equation in international trade: some microeconomic foundations and empirical evidence. Rev Econ Stat 67:474–481

Bodini A, Bondavalli C, Allesina S (2012) Cities as ecosystems: growth, development and implications for sustainability. Ecol Model 245: 185–198

Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. Science 323:892–895

Caberlotto L, Lauria M, Nguyen T-P, Scotti M (2013) The central role of AMP-kinase and energy homeostasis impairment in Alzheimer's disease: a multifactor network analysis. PLoS One 8:e78919

CEPII (2016) BACI World trade database:2016

Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Phys Rev E 70:66111

Csardi G, Nepusz T (2006) The igraph software package for complex network research

Dale BE, Kim S (2014) Can the predictions of consequential life cycle assessment Be tested in the real world? Comment on "using attributional life cycle assessment to estimate climate-change mitigation...". J Ind Ecol 18:466–467

De Benedictis L, Tajoli L (2011) The world trade network. World Econ 34:1417–1454

De Rosa M, Schmidt J, Brandão M, Pizzol M (2016) A flexible parametric model for a balanced account of forest carbon fluxes. Int J Life Cycle Assess. doi:10.1007/s11367-016-1148-z

Deng Y, Tian Y (2015) Assessing the environmental impact of flax fibre reinforced polymer composite from a consequential life cycle assessment perspective. Sustainability 7:11462–11483

Ekvall T, Weidema BP (2004) System boundaries and input data in consequential life cycle inventory analysis. Int J Life Cycle Assess 9: 161–171

Enders W (2014) Applied econometric time series, 4th edition. Wiley

Eshun JF, Potting J, Leemans R (2010) Inventory analysis of the timber industry in Ghana. Int J Life Cycle Assess 15:715–725

FAOSTAT (2016) Website of the food and agriculture organization of the United Nations. http://faostat.fao.org/

Fath BD, Scharler UM, Ulanowicz RE, Hannon B (2007) Ecological network analysis: network construction. Ecol Model 208:49–55

Fortunato S (2010) Community detection in graphs. Phys Rep 486:75–174

Grinde M (2011) Environmental assessment of scenarios for products and services based on forest resources in Norway. Institutt for energi- og prosessteknikk

Hänninen R, Hetemäki L, Hurmekoski E (2014) European forest industry and forest bioenergy outlook up to 2050: A synthesis
Hausmann R, Hidalgo CA (2011) The network structure of economic output. J Econ Growth 16:309–342
Heijungs R (2012) Spatial differentiation, GIS-based regionalization, hyperregionalization, and the boundaries of LCA. In: Ioppolo G (ed) Environment and energy (editorial series of Italian commodity science academy and engineering Association of Messina). FrancoAngeli, Milano, pp. 165–176
Helin T, Sokka L, Soimakallio S et al (2012) Approaches for inclusion of forest carbon cycle in life cycle assessment—a review. GCB Bioenergy 5:475–486
Hetemäki L (2014) Future of the European forest-based sector: what science can tell us. Grano Oy
Hidalgo CA, Hausmann R (2009) The building blocks of economic complexity. Proc Natl Acad Sci U S A 106:10570–10575
Huang J, Ulanowicz RE (2014) Ecological network analysis for economic systems: growth and development and implications for sustainable development. PLoS One 9:e100923
Hurmekoski E (2016) Long-term outlook for wood construction in Europe. School of Forest Sciences, Faculty of Science and Forestry, University of Eastern Finland
Jørgensen S, Hauschild M (2013) Need for relevant timescales when crediting temporary carbon storage. Int J Life Cycle Assess 18: 747–754
Kim H, Holme P (2015) Network theory integrated life cycle assessment for an electric power system. 7:10961–10975
Levasseur A, Lesage P, Margni M et al (2010) Considering time in LCA: dynamic LCA and its application to global warming impact assessments. Environ Sci Technol 44:3169–3174
Lund H, Mathiesen B, Christensen P, Schmidt J (2010) Energy system analysis of marginal electricity supply in consequential LCA. Int J Life Cycle Assess 15:260–271
Mathiesen BV, Munster M, Fruergaard T et al (2009) Uncertainties related to the identification of the marginal energy technology in consequential life cycle assessments. J Clean Prod 17:1331–1338
Navarrete-Gutiérrez T, Rugani B, Pigné Y et al (2015) On the complexity of life cycle inventory networks: role of life cycle processes with network analysis. J Ind Ecol 20:1094–1107
Neupane B, Halog A, Dhungel S (2011) Attributional life cycle assessment of woodchips for bioethanol production. J Clean Prod 19:733–741
Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69:26113
Nguyen T-P, Scotti M, Morine MJ, Priami C (2011) Model-based clustering reveals vitamin D dependent multi-centrality hubs in a network of vitamin-related proteins model-based clustering reveals vitamin D dependent multi-centrality hubs in a network of vitamin-related proteins. BMC Syst Biol 5:1752–1509
Nuss P, Chen W-Q, Ohno H, Graedel TE (2016) Structural investigation of aluminum in the U.S. economy using network analysis. Environ Sci Technol 50:4091–4101
Pizzol M, Scotti M, Thomsen M (2013) Network analysis as a tool for assessing environmental sustainability: applying the ecosystem perspective to a Danish water management system. J Environ Manag 118:21–31
R Core Team (2005) R: a language and environment for statistical computing
Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. Phys Rev E. doi:10.1103/PhysRevE.74.016110
Reichardt J, Bornholdt S (2007) Clustering of sparse data via network communities—a prototype study of a large online market. Journal of Statistical Mechanics: An IOP and SISSA Journal. doi:10.1088/1742-5468/2007/06/P06016
Reinhard J, Weidema B, Schmidt J (2010) Identifying the marginal supply of wood pulp. 2.-0 LCA Consultants, Aalborg
Rodriguez MA, Pepe A (2008) On the relationship between the structural and socioacademic communities of a coauthorship network. J Informetr 2:195–201
Schaubroeck T, Staelens J, Verheyen K et al (2012) Improved ecological network analysis for environmental sustainability assessment; a case study on a forest ecosystem. Ecol Model 247:144–156
Schaubroeck T, Alvarenga RAF, Verheyen K et al (2013) Quantifying the environmental impact of an integrated human/industrial-natural system using life cycle assessment; a case study on a forest and wood processing chain. Environ Sci Technol 47:13578–13586
Schmidt JH (2010) Comparative life cycle assessment of rapeseed oil and palm oil. Int J Life Cycle Assess 15:183–197
Schmidt JH (2015) Life cycle assessment of five vegetable oils. J Clean Prod 87:130–138
Scott-Boyer MP, Lacroix S, Scotti M et al (2016) A network analysis of cofactor-protein interactions for analyzing associations between human nutrition and diseases. Sci Rep 6:19633
Singh S, Bakshi BR (2011) Insights into sustainability from complexity analysis of life cycle networks: a case study on gasoline and bio-fuel networks. Proceedings of the 2011 I.E. International Symposium on Sustainable Systems and Technology
Tukker A, de Koning A, Wood R et al (2013) EXIOPOL—development and illustrative analyses of a detailed global MR EE SUT/IOT. Econ Syst Res 25:50–70
Wasserman S, Faust K (2016) Social network analysis—methods and applications. Cambridge University Press, Cambridge
Weidema B, Frees N, Nielsen A-M (1999) Marginal production technologies for life cycle inventories. Int J Life Cycle Assess 4:48–56
Weidema B, Ekvall T, Heijungs R (2009) Guidelines for application of deepened and broadened LCA—Deliverable D18 of work package 5 of the CALCAS project. ENEA, The Italian National Agency on new Technologies, Energy and the Environment
Wood R, Stadler K, Bulavskaya T et al (2015) Global sustainability accounting—developing EXIOBASE for multi-regional footprint analysis. Sustainability 7:138
Zamagni A, Guinée J, Heijungs R et al (2012) Lights and shadows in consequential LCA. Int J Life Cycle Assess 17:904–918