International migration network: Topology and modeling

Giorgio Fagiolo
LEM, Sant'Anna School of Advanced Studies, Pisa, 56127 Italy

Marina Mastrorillo
Princeton University, Princeton, New Jersey 08540, USA

(Received 29 January 2013; revised manuscript received 5 June 2013; published 18 July 2013)

This paper studies international migration from a complex-network perspective. We define the international migration network (IMN) as the weighted-directed graph where nodes are world countries and links account for the stock of migrants originated in a given country and living in another country at a given point in time. We characterize the binary and weighted architecture of the network and its evolution over time in the period 1960–2000. We find that the IMN is organized around a modular structure with a small-world binary pattern displaying disassortativity and high clustering, with power-law distributed weighted-network statistics. We also show that a parsimonious gravity model of migration can account for most of observed IMN topological structure. Overall, our results suggest that socioeconomic, geographical, and political factors are more important than local-network properties in shaping the structure of the IMN.

DOI: 10.1103/PhysRevE.88.012812 PACS number(s): 89.65.Gh, 89.70.Cf, 89.75.−k, 02.70.Rr

I. INTRODUCTION

This paper studies the evolution of human international migration patterns in the period 1960–2000 using a complex-network perspective [1–3]. International migration is increasingly perceived as a fundamental feature of human life for its huge impacts on the global economy and the potential to shape the world where we live by changing the demographic structure of towns, cities, and nations [4–6]. Recent predictions suggest that international migration is likely to become bigger and more complex in the coming decades, due to the combined impact of population aging and increasing demographic differences, economic inequality, climate change, environmental disasters, wars and famines, new political and economic scenarios, and technological change favoring social contacts [7].

So far, the lack of coordinated and homogeneous data on worldwide migration has forced researchers to focus either on country-specific data or to address global migration issues using very small samples of reporting countries [8,9]. In all these studies [10], international migration has been considered as a bilateral phenomenon, i.e., migration between any two countries was taken as independent of indirect links with seemingly unrelated partners [11]. More recently, however, a reliable and homogeneous bilateral-migration database covering most of world countries over the period 1960–2000 has been made available [12]. This eventually allows one to investigate the evolution of international migration patterns from a complex-network perspective. Building on this idea, this paper studies the international-migration network (IMN), defined as the weighted-directed human-migration graph where nodes represent world countries and directed (weighted) links describe international migration corridors (and related migration intensities) between any two countries. This parallels what has been recently done for international-trade [13,14] and financial [15] macroeconomic networks.

Despite international migration data being naturally described as a network, only a few studies have tried to recast this phenomenon using a graph-theoretic perspective [16–18]. A network perspective has been instead heavily used to study human-mobility issues in general [19–21]; air, cargo, and maritime transportation networks [22–25]; and interurban traffic [26]. These works show that understanding human-mobility networks can be important to predict the properties of diffusion processes occurring on them, e.g., global epidemics [27,28].

The rest of the paper is organized as follows. Section II defines the international-migration network and discusses its topological properties. Section III studies two different classes of models that can be employed to explain IMN topology. Finally, Sec. IV concludes.

II. TOPOLOGICAL PROPERTIES OF THE INTERNATIONAL-MIGRATION NETWORK

We employ migration data from the United Nations Global Migration Database [12] to build a time sequence of weighted-directed networks describing bilateral migration stocks [29]. More precisely, we define the weighted IMN at time $t = 1960, \ldots, 2000$ as the (directed) network characterized by the $N \times N$ weight matrix M^t, whose generic element $m^t_{i,j}$ represents the stock of migrants originated in country i and present at time t in country j. Accordingly, we define the binary projection of the IMN through the sequence of adjacency matrices $A^t = \{a^t_{i,j}\}$, where $a^t_{i,j} = 1$ iff $m^t_{i,j} > 0$, and zero otherwise. Note that since, in general, $m^t_{i,j} \neq m^t_{j,i}$, $m^t_{i,j} > 0$ does not necessarily imply $m^t_{j,i} > 0$, both weighted and binary IMNs are directed (nonsymmetric) networks. Figure 1 plots the undirected weighted version of the IMN in year $t = 2000$. In the figure, link directions are suppressed to attain a better visualization of the graph and link thickness is proportional to the logs of total bilateral migrants $(m^t_{i,j} + m^t_{j,i})$. To get a feel of migration determinants, node size is made proportional to logged country population, while node color (from beige to red, i.e., from lighter to darker grey) represents country income, measured by country per-capita gross domestic product (pcGDP). Only links associated to a

 ©2013 American Physical Society

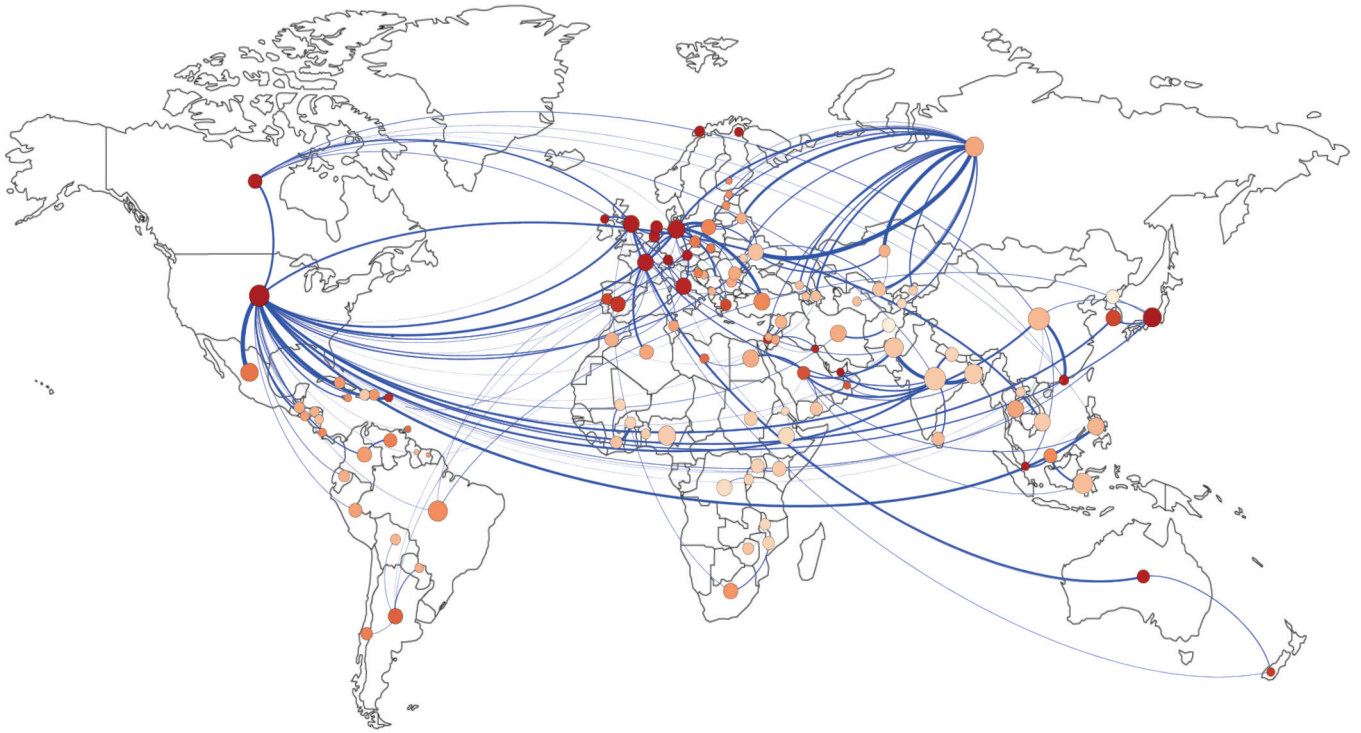

FIG. 1. (Color online) The IMN in year 2000. The figure plots the undirected weighted version of the IMN where only bilateral link weights (total number of bilateral migrants) larger than 200 000 are reported. The thickness of links in the plot is proportional to the logs of link weights. The size of the nodes is proportional to the log of country population. Node color represents country income, measured by country pcGDP. Red (darker grey), higher country pcGDP.

number of total bilateral migrants larger than 200 000 are shown. The map allows one to appreciate the central role of the United States and Russia, the importance of Gulf and Asian countries, the widespread presence of low-income countries, and the local nature of many migration corridors.

A. Binary and weighted connectivity

A thorough analysis of the main structural properties of the IMN (see Sec. II, Supplemental Material [29]) reveals, to begin with, that the volume of the network (total number of migrants) almost doubled from 1960 to 2000, from 93 million to 167 million. Network density also increased, from 0.32 to 0.47, thanks to the creation of new migration corridors. As a result the IMN is almost fully strongly connected, with its giant component displaying a small constant diameter and a decreasing average (directed) path length (APL). This suggests that the increase in density and number of reciprocated links over time, due to historical trends discussed above, has resulted in shortening the overall distances between countries in the IMN, either by creating directed corridors between countries previously not connected, or by forming shortcuts able to bypass longer paths.

The net increase in terms of average link weight was also positive, from more than 5000 to 7000. However, link-weight distribution appears to be extremely skewed. For this reason, we use logs of migrant stock to weigh the links when computing weighted-network statistics in what follows. Indeed, as Fig. 2 shows, link weights are power-law distributed across all the years with a stable power-law exponent of about 1.3. This strongly differs from log-normal weights that one typically finds in the ITN and other macroeconomic networks [13,15] and it is well in tune with the push-pull literature on migration determinants [9,30]. According to this theory, migration occurs if the reason to emigrate (the push) is balanced by a corresponding pull at destination. Push factors incentivizing emigration

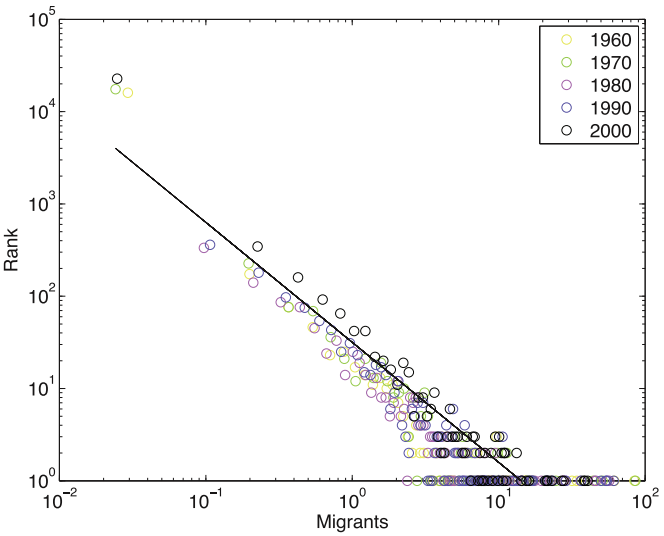

FIG. 2. (Color online) Log-log plot of the link-weight distribution (number of migrants) in different years. Link weights are rescaled by the volume of the IMN in order to wash away the trend. Power-law exponent: 1.36.

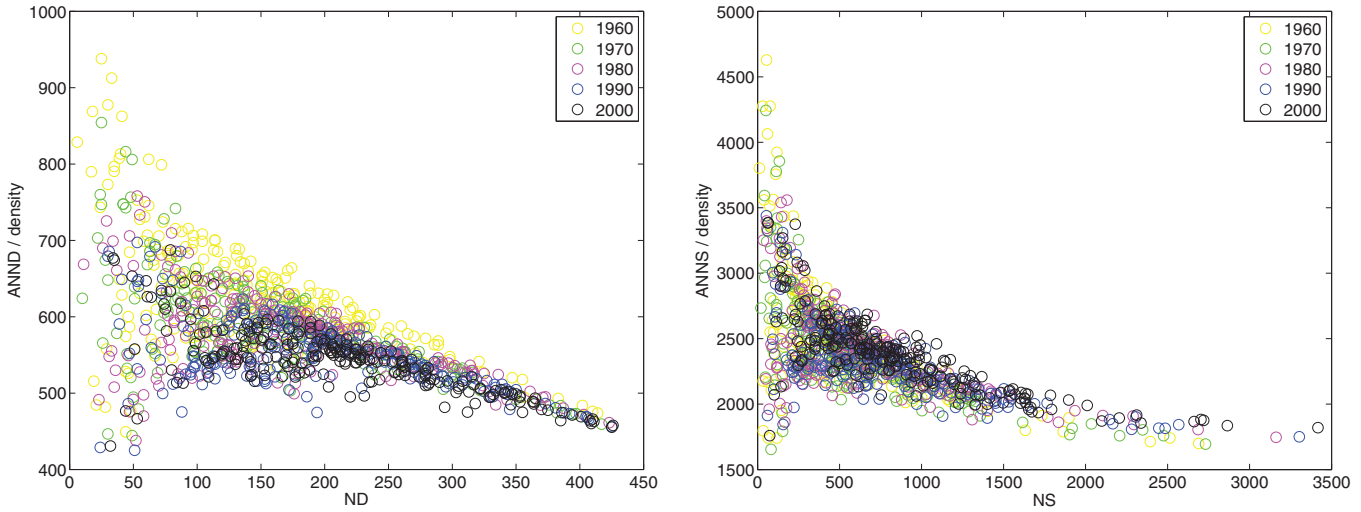

FIG. 3. (Color online) Disassortative behavior in the weighted IMN. Disassortative behavior in the binary IMN. Left: ANND rescaled by network density vs total node degree in different years. Right: ANNS rescaled by network density vs total node strength in different years.

include the lack of economic opportunities, religious or political persecution, hazardous environmental conditions, and so on. Pull factors at destination instead comprise availability of jobs, religious or political freedom, and the perception of a relatively benign environment. Existence of institutions at the origin and well-established communities of foreign-born people at destination also fosters migration. If pull and push factors induce persistent gaps between origin and destination along their development phases, they are likely to trigger a self-enforcing process where strong migration corridors increasingly attract people. Power-law behavior also characterizes node total immigrants and emigrants, i.e, in- and out-node strength (NS) (see Supplemental Material [29] and Ref. [31] for definitions). Remarkably, power laws emerge also in the distribution of the shares of total immigrant and emigrant to country population, meaning that fat tails are not the effect of heterogeneous country size.

As expected, one does not observe a similar skewness in node degree (ND) distributions, which typically exhibit a characteristic scale. They also display a high level of stability across the years, due to the high persistence of link formation and deletion process: Over two consecutive decades, the probability of link formation grew from 7.8% to 12.8%, whereas the likelihood that an already established corridor is severed went from 6.5% to 8%. This means that between any two consecutive decades the IMN binary architecture has remained quite stable: On average 91.4% of country pairs did not change their status (linked or not linked). These figures closely match those of the ITN [13]. We also find that the number of immigration and emigration corridors is identically distributed (this is confirmed also by two-sample Kolmogorov-Smirnov statistical tests) and unimodal across the years, which implies that, despite the fact that countries do not reciprocate all the directed bilateral links, they keep on average an identical number of inward and outward migration links.

B. Assortativity, clustering, and path length

An important feature of complex networks is their assortative or disassortative pattern [32,33]. We are interested in investigating if better (or more-intensively) connected nodes are in turn linked with better (or more-intensively) connected partners. We study this by computing Pearson's correlation between ND (respectively, NS) and average node nearest-neighbor degree (ANND) and strength (ANNS), i.e., population-average degree (strength) of the neighbors of a given country. As Fig. 3 shows, both binary and weighted versions of the IMN are (weakly) disassortative. Large players in the IMN are typically connected to countries that hold on average a small number of migration corridors or channel a small stock of migrants. Notice that instead small players are very heterogeneous. Mostly due to geopolitical constraints, some small countries, typically Pacific islands, are locally connected to few other geographically close countries that are themselves poorly connected. Other small countries instead hold privileged corridors to big players (e.g., European and North American countries) and therefore display a disassortative mixing.

In many networks, disassortative patterns are associated to small levels of binary or weighted clustering [34,35]. We find instead that the IMN displays a very high level of average clustering. Average binary clustering coefficient (BCC), i.e., population average of the probability that any two partners of a node are themselves partners, increases from 0.64 to 0.69 in our sample year. Correspondingly, average weighted clustering coefficient (WCC), which measures local clustering by weighting each link of a triangle by the correspondent total migration stock, goes from 0.13 to 0.15, a relatively high value, comparable to that in the ITN when logs of export flows are employed to weigh the links. We also find that node BCC is negatively correlated with both ND and NS (see Fig. 4 for an example). This hints at a hierarchical structure where very connected nodes, in terms of both ND and NS, typically form few triangles with their neighbors—a phenomenon possibly driven by the few large hubs present in the IMN. We also find that nodes that either hold a small number of migration corridors, or a small number of migrants, or both, are typically connected with pairs of countries that are also connected. However, the intensity of these triangles is the same as those

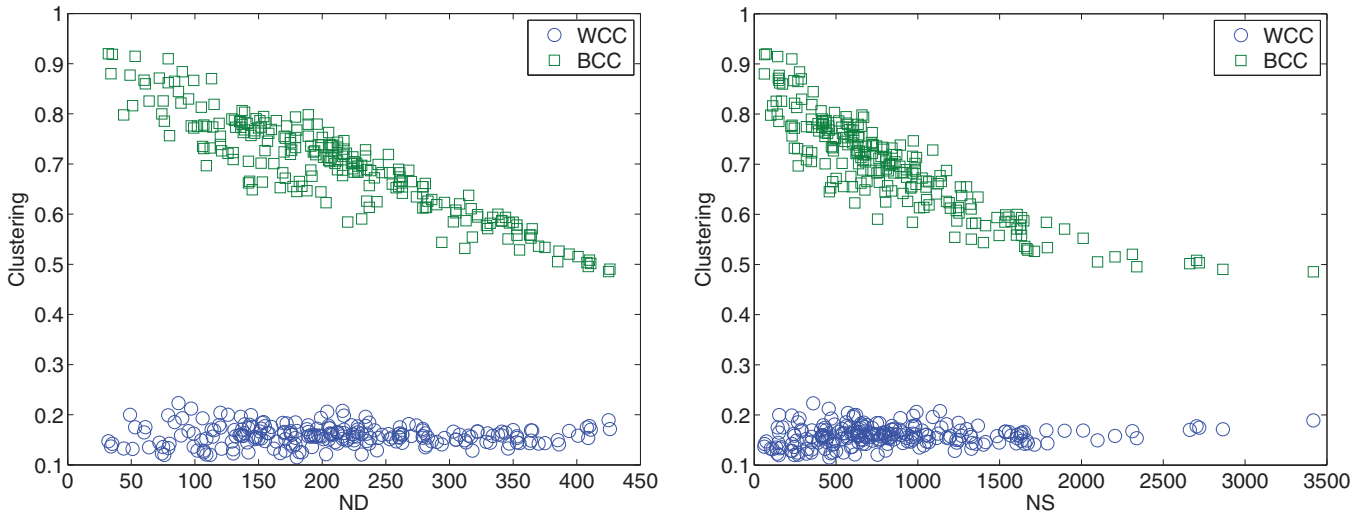

FIG. 4. (Color online) Left: Node clustering coefficients vs ND in year 2000. WCC, weighted clustering coefficient. BCC, binary clustering coefficient. Right: Node clustering coefficients vs NS in year 2000. WCC, weighted clustering coefficient. BCC, binary clustering coefficient.

created by nodes holding a lot of intense connections i.e., most of the existing triplets are really of a weak intensity.

Small and decreasing levels of APL, together with high binary clustering may be symptoms of small-world behavior for the binary IMN [36]. Figure 5 shows that this is actually the case. We plot global binary clustering coefficient (CC) [37] and (undirected) APL against their expected values in corresponding Erdős-Rényi (ER) random graphs where observed density is preserved. Note that APL is larger than its expected value and decreases over time, whereas the clustering coefficient is 30% larger than its expected value in ER graphs and keeps increasing over time from 0.67 to 0.76. This means that, as we approach year 2000, the formation of new links has gradually filled previously uncompleted triplets, thus contributing to the decrease of geodesic distances and to the increase in clustering. A larger number of previously unconnected countries have

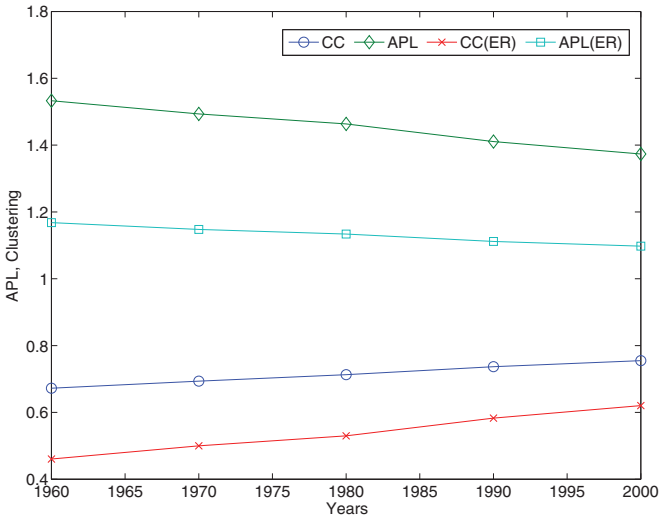

FIG. 5. (Color online) Small-world behavior of the IMN. We plot network-wide binary clustering coefficient [36] and average-path length in the binary undirected IMN together with their values in random ER graphs preserving observed IMN density in each year.

opened migration corridors between them. This has increased the probability that any two countries that shared a migration corridor with a target country became themselves migration partners.

C. Rich club, core-periphery, and community structure

The existence of a negative correlation between connectivity and binary clustering, and the absence of such association for weighted clustering, coupled with binary and weighted disassortativity, points to an overall topological structure for the IMN where concentration of strong migration corridors between a small number of rich-club countries is quite unlikely. As discussed in the Supplemental Material [29], this is confirmed by more formal tests, which show that the IMN does not feature either a binary or a weighted rich-club feature over the years.

All this in turn suggests that the IMN is a relatively poorly concentrated network, with a binary small-world structure where a few hubs and a lot of local structures coexist, generating high clustering and disassortativity. To formally check this intuition, we have performed a weighted community-structure analysis of IMN, using a weighted version of the standard community-detection Newman-Girvan modularity algorithm [38,39]. We find that the IMN displays relatively high values of optimized modularity (Q) and features, as expected, quite a rich structure of clusters (see Supplemental Material [29] for details). The number of communities decreases across time (from 14 to 7), suggesting that globalization has made the architecture of the IMN less fragmented and modules more strongly interconnected between them. Formal comparison tests made using normalized mutual information (NMI, see [40]) show that the community structure of the IMN has remained relatively stable until 1990, and then underwent some reorganization in the last decade of our sample.

IMN community structure is plotted in Fig. 6, which compares world maps in 1960 and 2000 (countries are colored according to the community they belong to). In 1960,

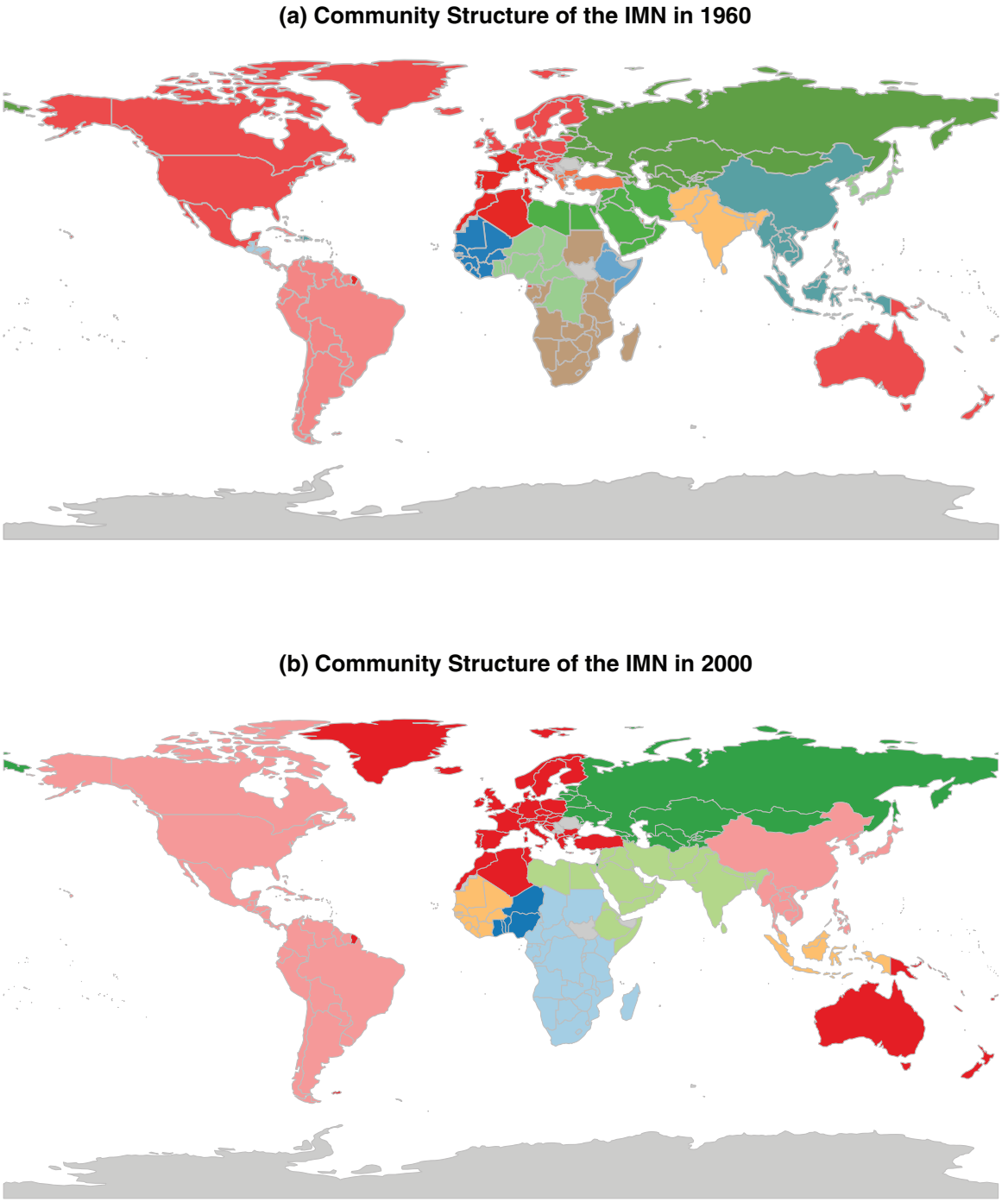

FIG. 6. (Color online) Community detection in the weighted IMN. World maps show in 1960 (top) and 2000 (bottom) communities detected using the Newman-Girvan modularity algorithm [38] using Tabu Search [54]. Countries belonging to same community are plotted in the same color/grayscale value.

IMN community structure was much more geographically fragmented. South and North American countries belonged to different communities. Europe was also separated in two clusters (Italy, France, Spain, and Portugal vs central and eastern countries). Turkey, Greece and Rumania were also forming a separate group, as it happened for eastern asian countries and the Indian block. In 2000, there existed a unique big American cluster. Another big cluster ranging from India to Libya and involving Gulf states was formed. European countries became integrated (including Turkey), and so did central and southern African states. Interestingly, a cluster of countries gravitating around Indonesia formed, probably due to the increasing economic role of that country in the Asian scenario. Note that, as expected, the role of geographical constraints emerges very clearly: Migration clusters are strongly shaped by regional and proximity factors. To better understand the extent to which geographical, demographical, and economic factors explain the structure of communities in the IMN, we have computed

the NMI index between observed IMN community structure and the partitions exogenously induced by a classification of countries according to geographical macro areas, population size, language, religion, income, gross-domestic product (GDP) levels and growth. As shown in greater detail in the Supplemental Material [29], and in accordance with our expectations, we find that the geography, population, and common language are the variables that induce partitions that are the most similar to migration-implied ones.

III. DETERMINANTS OF IMN STRUCTURE: NULL VS ECONOMETRIC MODELS

We now investigate whether the topological structure of the IMN, as summarized in Sec. II, can be explained by simple statistical or econometric models. We begin by fitting to the IMN null network models. By doing so, we want to understand if local-network statistics alone (node degrees and strength) can explain high-order properties of the network, e.g., disassortativity and clustering-connectivity correlation. Then we try to assess the extent to which a set of reasonable determinants of migrations (geographical distance, country income, language, religion, etc.), which are typically identified to be the most relevant drivers of migration stocks [9,12,41], can account for the observed structure of the IMN. As target variables to be explained, we focus in particular on node connectivity, ANND and ANNS, and clustering levels (BCC and WCC), and on assortative mixing and cluster-degree/strength correlation (leaving for a subsequent study the exploration of additional topological properties, e.g., motifs [42]).

A. Null models

In line with the recent literature on trade networks [43,44], we ask whether the observed IMN binary and weighted topology can be replicated by fixing its local structure only (degree and strength sequences) and leave everything else at random. More precisely, we employ the analytical method developed in Ref. [45] to compute expected statistics (and confidence intervals) for higher-order network statistics (assortativity, clustering) when one constrains both the in-degree and out-degree (or in- and out-strength) to be equal to the observed ones. Expected values are then computed over the grand-canonical ensemble of all possible random graphs obeying on average the constraints (see the Supplemental Material [29] for a more formal account).

Our main results are reported in Figs. 7–10, where we plot the evolution over time of correlation coefficients between expected and observed quantities in both the binary and the weighted (directed) IMN. Contrary to what happens in the ITN [43,44,46], local binary statistics (degrees) are not sufficient to control for higher-order statistics such as binary disassortativity and binary clustering. Furthermore, fixing in- and out-strength sequences is not enough to account for weighted disassortativity and clustering-strength correlation (which, as observed above, is almost always weak and not significant). The only topological feature that can be somewhat reproduced by the null model is the strong binary clustering-degree correlation, although from a statistical point of view expected correlation is significantly different from the

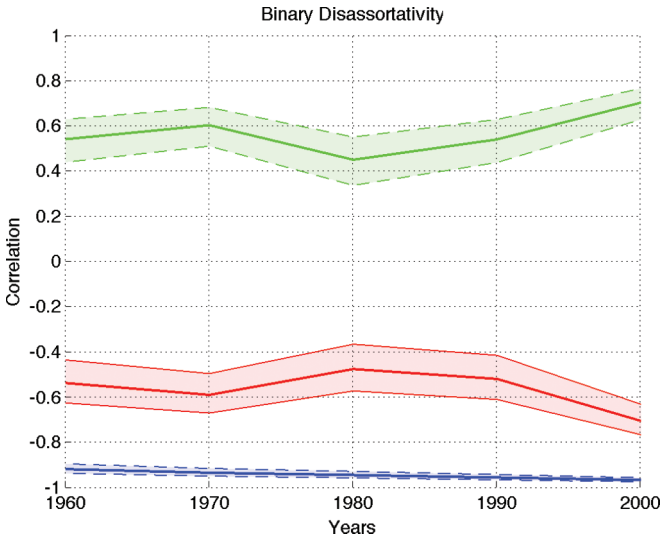

FIG. 7. (Color online) Null-model analysis of the binary directed IMN. Pearsons correlation coefficients between observed ANND vs observed ND [red (middle) line], expected ANND vs observed ND [blue (lower) line], and observed vs expected ANND [green (upper) line]. Expected network statistics computed fixing NDin and NDout sequences and applying the method in [45]. 95% confidence bands shown as shaded areas.

observed one in all the years. We interpret these results in the light of the underlying complexity of IMN structure. Despite the fact that some of the global structure like node centrality can be explained in terms of local node information, local network structure is not enough to explain most of the correlation structure between higher-order statistics and topological features. This seems to suggest that other determinants may be important to explain the topological structure of the IMN.

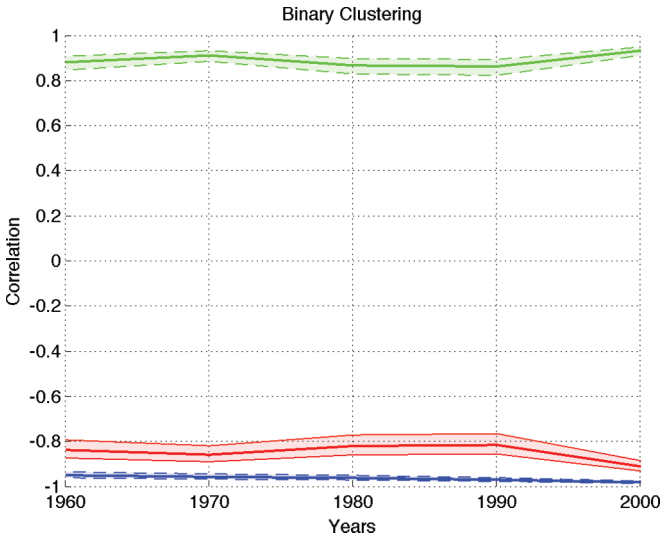

FIG. 8. (Color online) Null-model analysis of the binary directed IMN. Pearson correlation coefficients between observed BCC vs observed ND [red (middle) line], expected BCC vs observed ND [blue (lower) line], and expected vs observed BCC [green (upper) line]. Expected network statistics computed fixing NDin and NDout sequences and applying the method in [45]. 95% confidence bands shown as shaded areas.

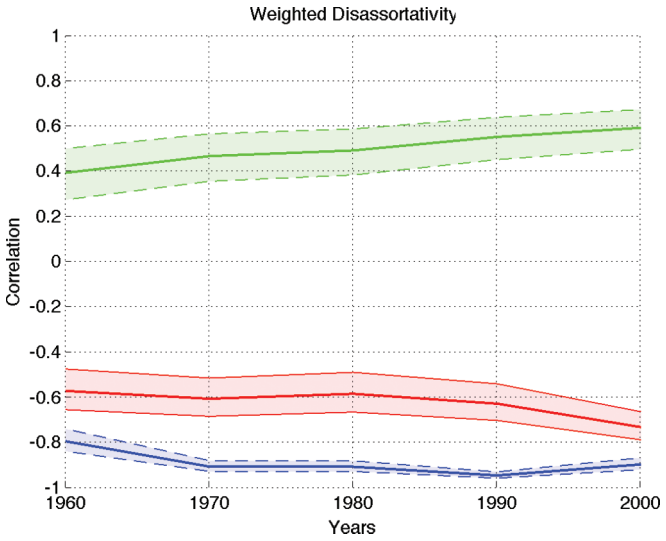

FIG. 9. (Color online) Null-model analysis of the weighted directed IMN. Pearsons correlation coefficients between observed ANNS vs observed NS [red (middle) line], expected ANNS vs observed NS [blue (lower) line], and observed vs expected ANNS [green/ (upper) line]. Expected network statistics computed fixing NSin and NSout sequences and applying the method in [45]. 95% confidence bands shown as shaded areas.

B. Gravity models and IMN architecture

The foregoing analysis suggests that knowledge about the local structure of the network is not enough to account for its higher-order properties and that other (non-network) determinants may be perhaps relevant. We address this issue by fitting bilateral migrant stock levels using a migration-based gravity model (GM) [47–50], which uses as explanatory factors

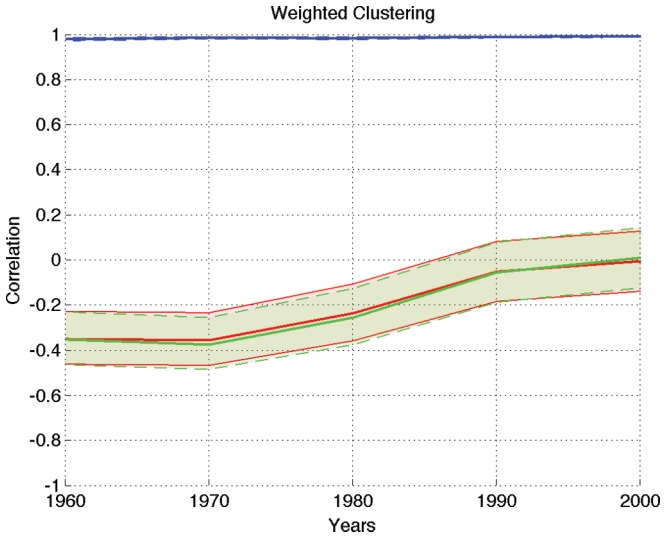

FIG. 10. (Color online) Null-model analysis of the weighted directed IMN. Pearsons correlation coefficients between observed WCC vs observed NS-tot [red (lower) line], expected WCC vs observed NS [blue (upper) line], and expected vs observed WCC [green (lower) line]. Expected network statistics computed fixing NSin and NSout sequences and applying the method in [45]. 95% confidence bands shown as shaded areas. Lower lines (green and red) and their confidence bands almost coincide.

country-specific economic and demographic variables (size, income, etc.) and bilateral country relationships (geographical distance, common language and religion, contiguity, etc.). We then employ the predictions of such a model to build a predicted IMN whose topological structure is compared to the observed one. More specifically, for any given year t in the sample we fit the following cross-section specification to the data (we omit time superscripts for simplicity):

$$m_{ij} = P_i^{\alpha_1} P_j^{\alpha_2} d_{ij}^{\alpha_3} rY_{ij}^{\alpha_4} \exp\{\beta_i + \beta_j + \boldsymbol{\gamma} \boldsymbol{Z}_{ij}\} \eta_{ij}, \qquad (1)$$

where P_i and P_j are country population, d_{ij} is geographical distance between i and j, rY_{ij} is relative per-capita GDP between i and j, η_{ij} is an error term uncorrelated with the independent variables, β_i and β_j are country dummies, and $\boldsymbol{Z}_{ij}$ is a vector of additional explanatory dummy variables controlling for (i) contiguity (i.e., border sharing), (ii) common official language, (iii) common religion, (iv) sharing of a colonial past relationship, (v) south-south (SS), south-north (SN), and north-north (NN) migration (e.g., $SS = 1$ if countries i and j are both in the south of the World). The parameters to be estimated are therefore $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_4)$, (β_i, β_j), and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_7)$.

We employ the model in Eq. (1) to estimate via a two-step procedure both the probability that any possible link ij is present ($p_{ij} = \text{Prob}\{m_{i,j} > 0\} = \text{Prob}\{a_{i,j} = 1\}$) and the expected value of the associated migrant stock $\hat{m}^*_{ij}$ (see the Supplemental Material [29] for more details). This generates two predicted probability-stock matrices ($\hat{P} = \{\hat{p}_{ij}\}$ and $\hat{M}^* = \{\hat{m}^*_{ij}\}$) that we use to simulate a large number (typically $H = 1000$) of predicted instances of both IMN adjacency and weight matrices, i.e., $(\hat{A}^h, \hat{M}^h)$, $h = 1, \ldots, H$. We perform two related exercises. First, we ask whether on average the predicted-binary matrices are similar to the observed one (A), in terms of both the percentage of successfully predicted zeros and ones and the topological structure of the binary IMN. Second, we investigate whether the weighted topological properties of the IMN can be successfully predicted by the gravity model using predicted instances of IMN adjacency and weight matrices.

For space reasons, we present results for year 2000 only (similar results hold in all the other years in the sample). Our first finding is that the gravity model provides a good explanation of the binary structure of the IMN. The logit specification is able on average to correctly predict the presence and the absence of a link in, respectively, 73.4% (z score = 7.674) and 82.3% (z score = 8.444) of cases.

This implies that on average the gravity model correctly reproduces both in-degree and out-degree observed IMN sequences [cf. Fig. 11(a)]. Both observed ANND and BCC are also well reproduced, albeit for large observed ANND and BCC values some underestimation occurs [cf. Figs. 11(c) and 11(d)]. Nevertheless, observed disassortativity and BCC-ND correlations are quite satisfactorily matched by the model [see Figs. 11(d) and 11(e)]. This result is somewhat surprising if one compares it with similar findings in the ITN [51], where instead a gravity specification mostly fails in reproducing the binary structure, while the knowledge of the local binary architecture is sufficient to reproduce high-order topology [46]. In the IMN, on the contrary, local topology is not enough

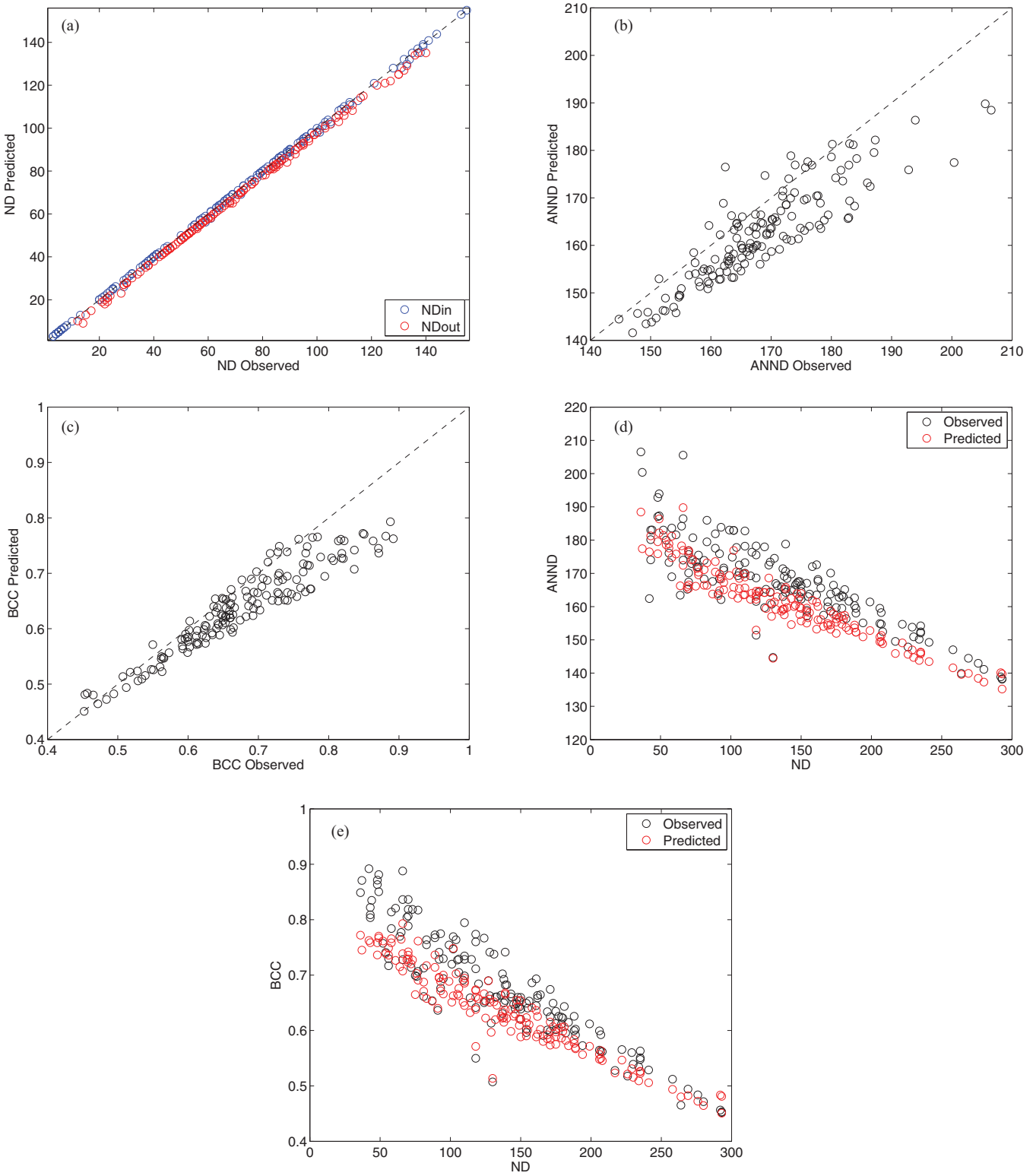

FIG. 11. (Color online) Gravity model estimation of the binary structure of the IMN. (a) Observed vs predicted node in degree (NDin) and out degree (NDout); NDin is depicted in blue and shows up slightly above NDout, in red. (b) Observed vs predicted ANND. (c) Observed vs predicted BCC. (d) Observed vs predicted correlation between ANND and total ND; observed values are depicted as black circles and are slightly above predicted ones (in red). (e) Observed vs predicted correlation between BCC and total ND; observed values are depicted as black circles and are slightly above predicted ones (in red). Each dot corresponds to the average across $H = 1000$ simulated binary structures using gravity-based logit predicted link probabilities.

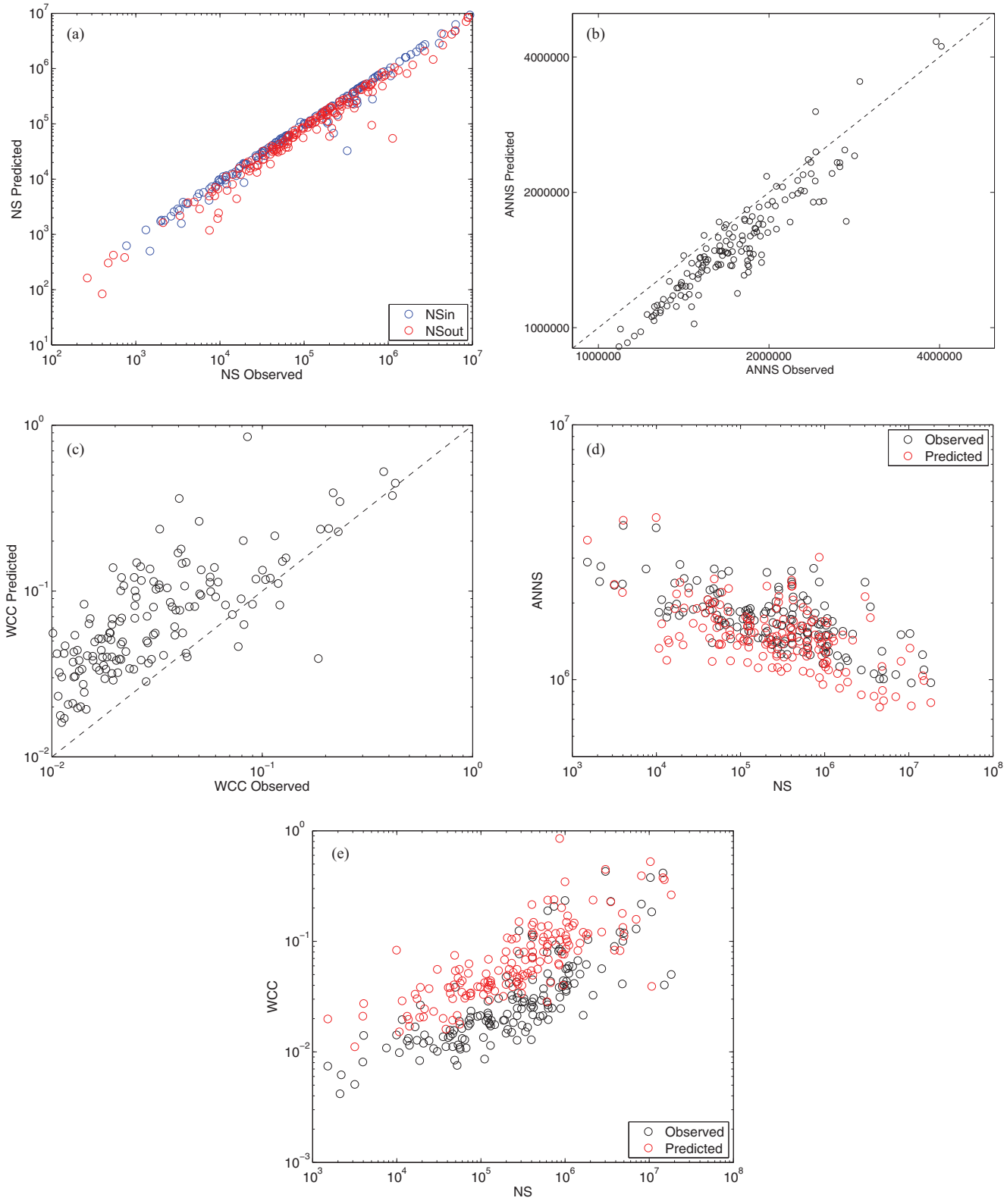

FIG. 12. (Color online) Gravity model estimation of the weighted structure of the IMN. (a) Observed vs predicted node in strength (NSin, blue circles) and out strength (NSout, red circles); NSin and NSout almost always lie on the main diagonal. (b) Observed vs predicted ANNS. (c) Observed vs predicted WCC. (d) Observed [black (darker)] vs predicted [red (lighter)] correlation between ANNS and total NS. (e) Observed [black (darker)] vs predicted [red (lighter)] correlation between WCC and total NS. Each dot corresponds to the average across $H = 1000$ simulated weighted IMN matrices binary structures. Each simulated matrix has a binary structure generated using gravity-based logit predicted link probabilities, on the top of which are superimposed zero-inflated poisson (ZIP) estimations of the relative migration stock (in levels) using gravity-based pseudomaximum likelihood predicted weights.

to replicate higher-order structure (see Sec. III A), whereas country-specific and bilateral country-interaction effects are able to account for most of the topological (binary) structure of migration corridors.

Motivated by the success of the gravity model in reproducing the observed binary topology of the IMN, we now turn to the analysis of its weighted structure. Our main findings are reported in Fig. 12, where we show observed vs gravity-predicted average weighted IMN statistics. All statistics are here computed on migrant levels (and not logs), as the gravity specification allows one to get a better precision, and then plotted in a log-log scale. Figure 12(a) shows that the gravity model is able to correctly predict in and out strength. This is not surprising, as strengths are linear combinations of bilateral migrant stocks, which are relatively nicely reproduced by the gravity model. Thanks to the good performance in replicating the binary structure in each simulation, also ANNS and WCC are well fitted by the model [Figs. 12(b) and 12(c)]. Finally, disassortativity is replicated as well as the positive correlation between weighted clustering and node strength (emerging when migrant stock levels and not their logs are used to weigh the links).

These results show that a relatively parsimonious specification of a gravity model of migration not only reproduces the IMN local network structure (degrees and strengths), but with some success also the underlying topology and the correlation structure between higher-order statistics. The goodness of fit of the migration-based gravity employed here is robust to alternative specifications, e.g., using additional variables sometimes employed in gravity studies (employment at origin and destination, institutional quality, etc.) [50] or alternative estimation techniques (panel data approach using time dummies, etc.).

Note that, on the contrary, in the ITN a gravity specification was able to correctly replicate the weighted structure only if the observed binary structure is kept fixed. Indeed, a trade-gravity model is not able to correctly predict the ITN binary topology [51]. In the IMN, instead, country-specific and bilateral interaction terms related to social, economic, geographical, and political factors can explain a good deal of details of the IMN. In particular, interacting terms controlling for origin-destination distance, geographical contiguity, and common sociopolitical features can greatly improve the explanation of country-specific variables that, as shown in the previous section, were only weakly significant in affecting local and higher-order node statistics. This has straightforward implications for both control and prediction of the future evolution of the IMN structure.

IV. CONCLUSIONS

International migration is increasingly recognized as a central issue to understand future trends in our globalized world. In this paper we have argued that describing migration data using a complex-network representation allows one to capture the complexity of international migration linkages between countries and gives one the possibility to study migration from a systemic perspective, where both direct and indirect linkages are taken into consideration. Our exercises show that the structure of the IMN has remained relatively stable through the years as far as migration corridors are concerned. However, the intensity of migration links has been strongly increasing leading to very skewed (power-law) distributions for weighted link and node statistics. The IMN is organized around quite a modular structure characterized by a binary small-world pattern displaying disassortativity and high clustering. We have also shown that geographical distance, country size, relative income, together with a host of country-specific and bilateral dummy variables including common language, common religion, and south-north linkages, are able to satisfactorily explain in a gravity-like fashion most of the binary and weighted topological structure of the IMN. Interestingly, the local topological structure of the network cannot fully replicate its higher-order structure. This points to a preponderance of socioeconomic, geographical, and political factors in the process shaping the IMN.

Our analysis can be extended in several ways. First, one may replicate the analysis in alternative link-weighting setups where migrants stocks are rescaled by country size measures (e.g., population), in order to check the robustness of our results in networks where one focuses on migration shares and not just levels. Secondly, one could try to fit to the data more sophisticated dynamic models of weighted network formation [52], where migrant corridors and flows are stochastically added over time, possibly as a function of network (weighted) topology. Finally, community-structure results can be strengthened and expanded both using different community-structure detection algorithms [53] and by modifying the null model employed within the modularity function. A possibility in the latter case is to use gravity-model predictions as expected values for link weights in the modularity function.

ACKNOWLEDGMENTS

Thanks to Matteo Chinazzi, Marco Dueñas, Giuseppe Mangioni, and Tiziano Squartini for their invaluable help in providing codes and assistance on network visualization, the econometrics of gravity exercises, community detection, and null models, respectively. G.F. gratefully acknowledges support received by the research project "The International Trade Network: Empirical Analyses and Theoretical Models" funded by the Italian Ministry of Education, University and Research (Scientific Research Programs of National Relevance 2009).

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford University Press, Oxford, 2007).
[3] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).
[4] Population Reference Bureau, Population Bulletin **63**, 1 (2008).

[5] *A Profile of Immigrant Populations in the 21st Century* (OECD Publishing, Paris, 2008).
[6] World Bank, *Global Economic Prospects* (World Bank, Washington DC, 2006).
[7] UNDP, *HDR 2009 - Overcoming Barriers: Human Mobility and Development* (United Nations, 2009).
[8] A. Dreher, N. Gaston, and P. Martens, *Measuring Globalisation: Gauging Its Consequences* (Springer, Berlin, 2008).
[9] N. Gaston and D. Nelson, in *Handbook of International Trade*, edited by D. Bernhofen, R. Falvey, D. Greenaway, and U. Kreickemeier (Palgrave, London, 2011).
[10] D. S. Massey, J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino, and E. J. Taylor, Popul. Dev. Rev. **19**, 431 (1993).
[11] G. Fagiolo, J. Econ. Interact. Coord. **5**, 1 (2010).
[12] C. Ozden, C. R. Parsons, M. Schiff, and T. L. Walmsley, World Bank Econ. Rev. **25**, 12 (2011).
[13] G. Fagiolo, J. Reyes, and S. Schiavo, Phys. Rev. E **79**, 036115 (2009).
[14] M. Barigozzi, G. Fagiolo, and D. Garlaschelli, Phys. Rev. E **81**, 046104 (2010).
[15] S. Schiavo, J. Reyes, and G. Fagiolo, Quant. Finance **10**, 389 (2010).
[16] P. Slater, arXiv:0809.2768v3.
[17] E. Tranos, M. Gheasi, and P. Nijkamp, Tinbergen Institute Discussion Papers 12-123/VIII, Tinbergen Institute, 2012.
[18] K. Davis, P. D'Odorico, F. Laio, and L. Ridolfi, PLoS ONE **8**, e53723 (2013).
[19] D. Brockmann and F. Theis, IEEE Pervasive Comput. **7**, 28 (2008).
[20] F. Simini, M. C. Gonzalez, A. Maritan, and A.-L. Barabasi, Nature (London) **484**, 96 (2012).
[21] B. Csaji, A. Browet, V. Traag, J.-C. Delvenne, E. Huens, P. V. Dooren, Z. Smoreda, and V. D. Blondel, Physica A **392**, 1459 (2013).
[22] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004).
[23] Y. Hua and D. Zhu, Phisica A **388**, 2061 (2009).
[24] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius, J. R. Soc., Interface **7**, 1093 (2010).
[25] O. Woolley-Meza, C. Thiemann, D. Grady, J. Lee, H. Seebens, B. Blasius, and D. Brockmann, Eur. Phys. J. B **84**, 589 (2011).
[26] A. De Montis, M. Barthélemy, A. Chessa, and A. Vespignani, Environ. Plan. B: Plan. Des. **34**, 905 (2007).
[27] V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani, Proc. Natl. Acad. Sci. USA **103**, 2015 (2006).
[28] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani, Proc. Natl. Acad. Sci. USA **106**, 21484 (2009).
[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.88.012812 for (i) data description; (ii) methodologies employed throughout the paper; and (iii) additional analyses regarding the study of the topological properties of the international migration network and their time evolution.
[30] E. Lee, Demography **3**, 47 (1966).
[31] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Physica A **346**, 34 (2005).
[32] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[33] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[34] G. Fagiolo, S. Schiavo, and J. Reyes, Physica A **387**, 3868 (2008).
[35] G. Fagiolo, Phys. Rev. E **76**, 026107 (2007).
[36] D. Watts and S. Strogatz, Nature (London) **393**, 440 (1998).
[37] R. D. Luce and A. D. Perry, Psychometrika **14**, 95 (1949).
[38] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[39] A. Arenas, J. Duch, A. Fernandez, and S. Gomez, New J. Phys. **9**, 176 (2007).
[40] L. Danon, A. D. Guilera, J. Duch, and A. Arenas, J. Stat. Mech.: Theory Exp. (2005) P09008.
[41] International Organization for Migration, *World Migration Report 2010* (International Organization for Migration, Geneva, 2010).
[42] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[43] T. Squartini, G. Fagiolo, and D. Garlaschelli, Phys. Rev. E **84**, 046117 (2011).
[44] T. Squartini, G. Fagiolo, and D. Garlaschelli, Phys. Rev. E **84**, 046118 (2011).
[45] T. Squartini and D. Garlaschelli, New J. Phys. **13**, 083001 (2011).
[46] G. Fagiolo, T. Squartini, and D. Garlaschelli, J. Econ. Interact. Coord. **8**, 75 (2013).
[47] G. K. Zipf, Am. Sociol. Rev. **11**, 677 (1946).
[48] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, MA, 1949).
[49] J. E. Cohen, M. Roig, D. C. Reuman, and C. Gogwilt, Proc. Natl. Acad. Sci. USA **105**, 15269 (2008).
[50] E. Letouzé, M. Purser, F. R. Rodriguez, and M. Cummins, *Revisiting the Migration-Development Nexus: A Gravity Model Approach* (University Library of Munich, Germany, 2009).
[51] M. Dueñas and G. Fagiolo, J. Econ. Interact. Coord. **8**, 155 (2013).
[52] G. Caldarelli, A. Chessa, I. Crimaldi, and F. Pammolli, Phys. Rev. E **87**, 020106 (2013).
[53] S. Fortunato, Phys. Rep. **486**, 75 (2010).
[54] F. Glover and M. Laguna, *Tabu Search* (Kluwer Academic Publishers, Boston, 1997).