

Complex Networks and their Applications

Complex Networks and their Applications

Edited by

Hocine Cherifi

CAMBRIDGE
SCHOLARS

P U B L I S H I N G

Complex Networks and their Applications,
Edited by Hocine Cherifi

This book first published 2014

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2014 by Hocine Cherifi and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-5370-4, ISBN (13): 978-1-4438-5370-5

TABLE OF CONTENTS

Preface	vii
Chapter One.....	1
Disentangling Spatial and Non-spatial Effects in Real Complex Networks Tiziano Squartini, Francesco Picciolo, Franco Ruzzenenti, Riccardo Basosi and Diego Garlaschelli	
Chapter Two	39
Online and Offline Sociality: A Multidimensional Complex Network Approach Matteo Zignani, Sabrina Gaito and Gian Paolo Rossi	
Chapter Three	76
Multi-Ego-Centered Communities Maximilien Danisch, Jean-Loup Guillaume and Bénédicte Le Grand	
Chapter Four	112
Complex Networks in Scientometrics Adam Matusiak and Mikolaj Morzy	
Chapter Five	132
Rumor Dynamics and Inoculation of Nodes in Complex Networks Anurag Singh and Yatindra Nath Singh	
Chapter Six	174
Exploratory Network Analysis: Visualization and Interaction Sébastien Heymann and Bénédicte Le Grand	
Chapter Seven.....	212
The Composite Centrality Framework Andreas Joseph and Guanrong Chen	
Chapter Eight.....	245
Complex Networks and Epidemiology Marco Alberto Javarone and Giuliano Armano	

Chapter Nine.....	261
Building Social Networks in Online Chats with Users, Agents and Bots Vladimir Gligorijević, Milovan Šuvakov and Bosiljka Tadić	
Chapter Ten	285
Complex Networks and Web Services Chantal Cherifi	
Chapter Eleven	320
Non-Overlapping Community Detection Hocine Cherifi	
Contributors.....	354

PREFACE

Complex network theory is an emerging multidisciplinary field of research that is spreading to many disciplines such as physics, engineering, biology, sociology and economics. The common feature of many systems encountered in these different scientific fields is that they can be represented as a graph with the nodes representing a set of individual entities and the links standing for the interactions between these entities. Regardless of their physical nature, complex networks share some common structural properties that distinguish them from purely random graphs. Inspired by the study of real-world systems rather than by theory and fuelled by the availability of large datasets and computing power, research on complex networks is booming. The primary goal of this book is to provide an overview of the multiple aspects of this fast-growing research area. It contains eleven chapters presenting a wide spectrum of recent development with emphasis on theory and applications in the field. Although this book is a collection of independent studies, it represents a cohesive work that provides the reader with an up-to-date picture of the state of the field. Collectively, these contributions highlight the impact of complex network theory on a variety of scientific disciplines. This book does not solely reflect the opinion of the author. Instead, it expresses the views of 25 researchers working in well-known universities and research institutions throughout the world. The readers of this book are expected to be involved in a range of interdisciplinary studies. With this aim in mind, care was taken to make it as readable as possible to newcomers. I am honored to bring you this book, which was generated by the contributions and discussions held at the Workshop on Complex Networks and their Applications. I would like to thank the contributors of the different chapters for their constructive effort. I hope that “Complex Networks and their Applications” will be useful to a large audience of experts and graduate students and that it will stimulate important developments in this exciting research area.

Hocine Cherifi

CHAPTER ONE

DISENTANGLING SPATIAL AND NON-SPATIAL EFFECTS IN REAL NETWORKS

TIZIANO SQUARTINI, FRANCESCO PICCIOLO,
FRANCO RUZZENENTI, RICCARDO BASOSI
AND DIEGO GARLASCHELLI

Over the last fifteen years, Network Science has facilitated the identification of universal and unexpected patterns across systems belonging to deeply different research fields, such as biology, economics and physics (Caldarelli 2007). A fruitful cross-fertilization among these disciplines, leading to the introduction of novel multidisciplinary tools, has been made possible by the fact that many real complex systems can be formally abstracted as networks or *graphs*, irrespective of their specific nature. In so doing, several details of the original system are discarded and the emphasis is put on the study of the topological properties of the underlying ‘network backbone’ (Caldarelli 2007; Barrat Barthelemy and Vespignani 2008; West Brown and Enquist 1997, 1999, 2001). While this process facilitates the detection of key structural properties in real complex systems, it can also obscure other important levels of organization that involve non-topological factors. A key example is the *spatial* organization of networks (Barthelemy 2011).

Many real networks lie embedded in a metric space, i.e. a space where *distances* between nodes can be properly defined. In such cases, besides their connectivity, vertices can be identified by additional parameters, definable as *coordinates*, measuring their position and allowing the quantification of their mutual “proximity”. We will refer to these networks as *embedded networks*. Embedded networks represent an important subset of real networks: transportation systems, electric power grids, wireless communication networks and the Internet (i.e. the net of physical connections between servers) are only a few examples of systems embedded

in a two-dimensional metric space (Barthelemy 2003, 2011; Emmerich et al. 2012; Woolley-Meza et al. 2011).

Social networks, e.g. those represented by friendship or sexual relations among individuals, are also shaped by the proximity of the nodes in a two-dimensional space (even if the World Wide Web is challenging our traditional way of establishing social relations, it is still far more common to have a higher number of friends in the same city or country than in a distant one). Other examples, such as neural networks and protein networks, can instead be considered as occupying a three-dimensional metric space (Emmerich et al. 2012).

The range of applications can be even further extended to networks that are not necessarily embedded in a physical or geographic space, by noticing that the concept of metrics allows us to study configurations lying in abstract (e.g. cultural, economic or temporal) spaces, where distances are defined accordingly (Axelrod 1997; Aiello et al. 2012; Starnini et al. 2012; Valori et al. 2012). For instance, networks of protein configurations linked by saddle-points in a properly defined energy landscape are examples of networks embedded in high-dimensional configuration spaces. In all these examples, both vertex-specific and global spatial dependencies affect the dynamics of the network (Böde 2007). Thus, in order to deepen our understanding of the mechanisms shaping real networks and ruling their evolution, the unavoidable step to be made is to take into account also spatial properties (Bettencour et al. 2007; Bejan and Lorente 2010; Emmerich et al. 2012).

Unfortunately, while many theoretical models have already been introduced in order to artificially generate networks shaped by a combination of spatial and non-spatial factors, it is still much more difficult to disentangle these two effects in real networks (Bradde et al. 2010; Barthelemy 2011; Picciolo et al. 2012).

Two main obstacles are encountered. First, most approaches require the introduction of a mathematical model where the functional dependence of network properties on distances is postulated *a priori* and thus arbitrarily (Duenas and Fagiolo 2011; Anderson and Yotov 2012). Second, it is very difficult to filter out a spurious or apparent component of spatial effects which is instead due to other non-spatial factors. For instance, hubs (vertices with many connections) are generally connected to several other nodes irrespective of the positions of the latter, simply because they are highly connected. This effect would generally appear as a local lack of spatial dependence, spuriously lowering any global measure of spatial effects, even if the overall network formation process were instead distance-driven. Conversely, pairs of hubs (vertices with many connections)

tend to contribute to an overestimation of spatial factors, since they are typically connected to each other even in networks where distances play no role. The distance between pairs of hubs would then incorrectly appear as a preferred spatial scale for connectivity, biasing again the interpretation of the results.

The above considerations clarify that, in order to disentangle spatial and non-spatial effects in real networks, any satisfactory approach should be able to control for two potentially misleading factors. First, it should control for the mathematical arbitrariness *a priori* associated to the definition of any proxy of spatial dependence. Second, it should control for the effects of non-spatial topological constraints inducing a spurious spatial dependence *a posteriori*, given the characteristics of the particular real network considered.

In this chapter, we describe how these two important prescriptions can be implemented into the definition of a general method that we have recently introduced (Ruzzenenti et al. 2012; Picciolo et al. 2012).

The method is based on the idea that, given any definition of spatial effects, the relevant information is not given by the measured value itself. A comparison is needed with the corresponding expected value under a suitable null model that preserves the non-spatial properties of the real network. This comparison removes the mathematical arbitrariness of the adopted definition, and the fact that the null model controls for non-spatial effects also removes the undesired effects of the latter. Moreover, by focusing on both global (network-wide) and local (vertex-specific) quantities, this method allows us to isolate the (potentially conflicting) contributions of individual nodes to the overall spatial effects.

We will describe our method in detail by considering its application to a particular embedded network, namely the World Trade Web (WTW) defined as the network of international import-export trade relationships between world countries. Our choice is driven by the fact that both spatial effects (e.g. geographic distances between countries) and non-spatial effects (e.g. the countries' Gross Domestic Products) are known to shape the structure of this network (Ruzzenenti et al. 2012; Picciolo et al. 2012). For this reason, the WTW is the ideal candidate not only to illustrate our method, but also to compare the results with a different class of spatial models known in the economic literature as *Gravity Models* (Tinbergen 1962; Linders Matijn and Van Oort 2008; Fagiolo 2010; Duenas and Fagiolo 2012; Squartini and Garlaschelli 2013).

As the name itself suggests, Gravity Models aim at predicting the yearly intensity of the total trade exchanges between any two countries by adopting the same functional form of Newton's gravitational potential.

The predicted intensity is proportional to the countries' GDPs (calculated in the same year as the trade exchanges) and inversely proportional to the countries' geographic distance (Tinbergen 1962; Linders Matijn and Van Oort 2008). Our results show that the effects of geographic distances on the WTW are much more complicated than what is generally learnt from the use of Gravity Models.

The remainder of the chapter is structured in three main sections and a final concluding one. In the first section, we consider the case of binary networks, where pairs of vertices are either connected or not connected. We initially define some preliminary quantities that will be our "target" measures of spatial effects. Then we introduce suitable null models where the role of distances is, in some sense, switched off. Finally we calculate some "integrated" quantities defined as combinations of observed and expected values. This allows us to assess whether spatial effects are present or not (both locally and globally), given a target quantity used as a proxy. In the second section, we extend our approach to weighted networks, where links can have different intensities. Taken together, the results of the first two sections reveal that spatial effects are clearly present in the WTW but vary considerably over time, for different countries, and for different (binary or weighted) representations of the network. In the third section, we extend our formalism in order to compare the entity of spatial effects with that of other factors shaping the network. We find that geographic distances are comparatively much less important than non-spatial properties such as the reciprocity of the network. We conclude that the role of distances in the WTW, in both absolute and relative terms, is very different from what is generally thought.

Spatial effects in binary networks

In order to disentangle spatial and non-spatial effects in real embedded networks, the first step is to define quantities which measure how a network "feels" its embedding space. For illustrative purposes, in **Fig. 1** we show two extreme ways in which this can happen. In both panels, nodes represent the capitals of the countries adhering to the European Union and distances between nodes are proportional to the geographic distances between the EU capital cities. In the top panel, links are established between the geographically closest pairs of countries, originating a "spatially polarized" (or shrunk) configuration. In the bottom panel, the same number of links is instead drawn between the most distant

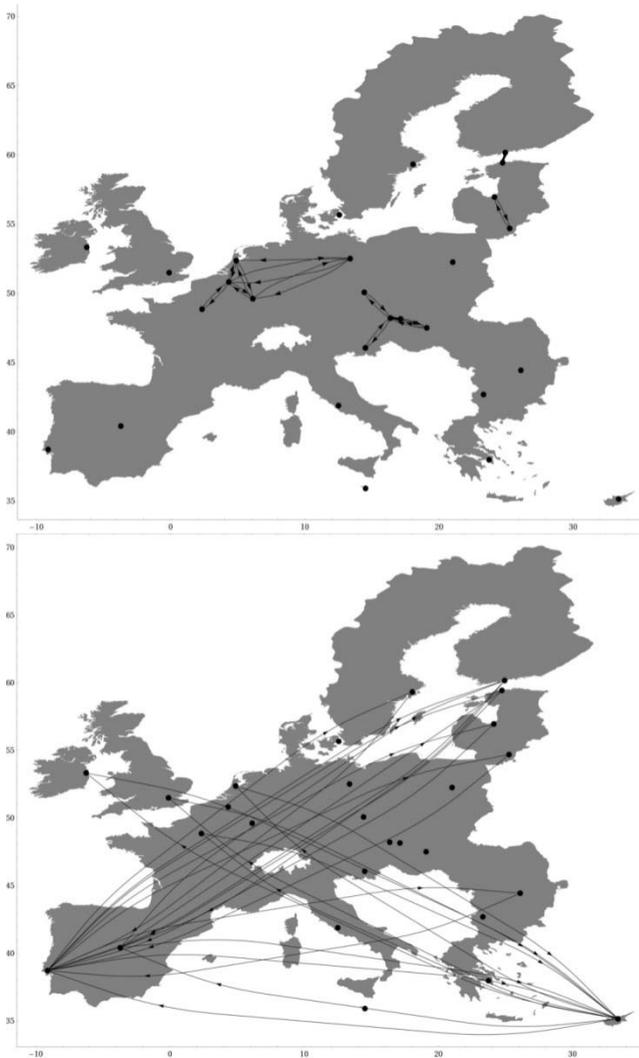


Fig. 1. Two examples of a hypothetical EU27 trading network ($N=27$). The black dots correspond to the geographic positions of the capital cities. For a given number of links (here arbitrarily chosen to be $L=27$), the figure represents the maximally shrunk network (top) and the maximally stretched network (bottom). The filling coefficient f that we introduce later takes the values $f = 0$ (top) and $f = 1$ (bottom) for these two extreme configurations, and $0 < f < 1$ in intermediate cases.

pairs of countries, originating a “spatially diluted” (or stretched) configuration. In the following subsections, we define quantities that can properly distinguish between these extremes and also capture any intermediate configuration.

A global measure

A binary, directed graph is specified by a $N \times N$ adjacency matrix, A , where N is the number of nodes and the generic entry a_{ij} is 1 when there is a connection from node i to node j , and 0 otherwise. The simplest definition of a global measure incorporating distances and network structure is

$$F = \sum_{i=1}^N \sum_{j \neq i} a_{ij} d_{ij} \quad (1)$$

where d_{ij} is the generic entry of the matrix of distances, D , among nodes (Ruzzenenti et al. 2012). Since we will consider networks without self-loops (i.e. $a_{ii} = 0$), F is a measure of the total distance between different, topologically connected pairs of nodes. Equivalently F can be seen as a measure of the extent to which the networks “fills” the available space.

The quantity F reaches its minimum when the links are placed between the closest vertices. Formally speaking, if we consider the list $V^\uparrow = (d_1^\uparrow, \dots, d_n^\uparrow, \dots, d_{N(N-1)}^\uparrow)$ of all non-diagonal elements of D ordered from the smallest to the largest ($d_n^\uparrow \leq d_{n+1}^\uparrow$), the minimum value of F is simply given by $F_{\min} = \sum_{n=1}^L d_n^\uparrow$, where $L = \sum_{i=1}^N \sum_{j=1}^N a_{ij}$ is the number of links in the network. Similarly, the maximum value of F is reached when links are placed between the spatially farthest nodes. Considering the list $V^\downarrow = (d_1^\downarrow, \dots, d_n^\downarrow, \dots, d_{N(N-1)}^\downarrow)$ of distances in decreasing order ($d_n^\downarrow \geq d_{n+1}^\downarrow$), the maximum value of F for a network with L vertices is $F_{\max} = \sum_{n=1}^L d_n^\downarrow$.

In order to compare, and possibly rank, different networks according to their values of F , a normalized quantity should be used. An improved global definition, which we will denote as *filling coefficient*, is

$$f = \frac{F - F_{\min}}{F_{\max} - F_{\min}} = \frac{\sum_{i=1}^N \sum_{j \neq i} a_{ij} d_{ij} - F_{\min}}{F_{\max} - F_{\min}} \quad (2)$$

where $0 \leq f \leq 1$. For the maximally shrunk and maximally stretched configurations shown in **Fig. 1**, the filling coefficient takes the values $f = 0$ and $f = 1$ respectively. Depending on the chosen links' disposition, different choices of the two values F_{\min} and F_{\max} can be made. As an example, in **Fig. 2** we show how the extreme values of F (or equivalently f) can change if different or additional constraints, beside the total number of links, are enforced on the network topology (e.g. imposing only one outgoing link for each vertex). So, in principle, F_{\min} and F_{\max} can be arbitrarily tuned to fit the best scenario for the network under consideration.

In the next section we will present a general method to disentangle spatial and non-spatial effects concurring to shape embedded networks and we will compare them with the existing ones.

The filling allows us to quantify the tendency of embedded networks to fill the metric space they are in. However the interplay between spatial and non-spatial effects in shaping a network topology can be unambiguously quantified only after having defined a proper reference model with which to compare the observed value of F whose aim is to discount as much as possible spurious non-spatial effects letting the genuine distances-induced ones emerge (Squartini and Garlaschelli 2011).

The reference models we will define in what follows are probabilistic in nature and known as *null models*. The methodology underlying a null model prescribes to choose only a portion of the available observed network's information and test how effective it is in explaining the rest of the (unconstrained) topology (Shannon 1948; Jaynes 1957; Holland and Leinhardt 1975; Wasserman and Faust 1994; Maslov and Sneppen 2002; Park and Newman 2004; Garlaschelli and Loffredo 2008; Squartini and Garlaschelli 2011). The effectiveness of the chosen set of constraints will be also tested over time, by analyzing different temporal snapshots of the same network. In so doing the presence of statistically significant trends through time can be highlighted.

Non-spatial null models

As previously mentioned, the aim of this comparison is to discount apparent or spurious spatial effects due to non-spatial factors. For this reason, we need to introduce space-neutral models that play the role of *null models*.

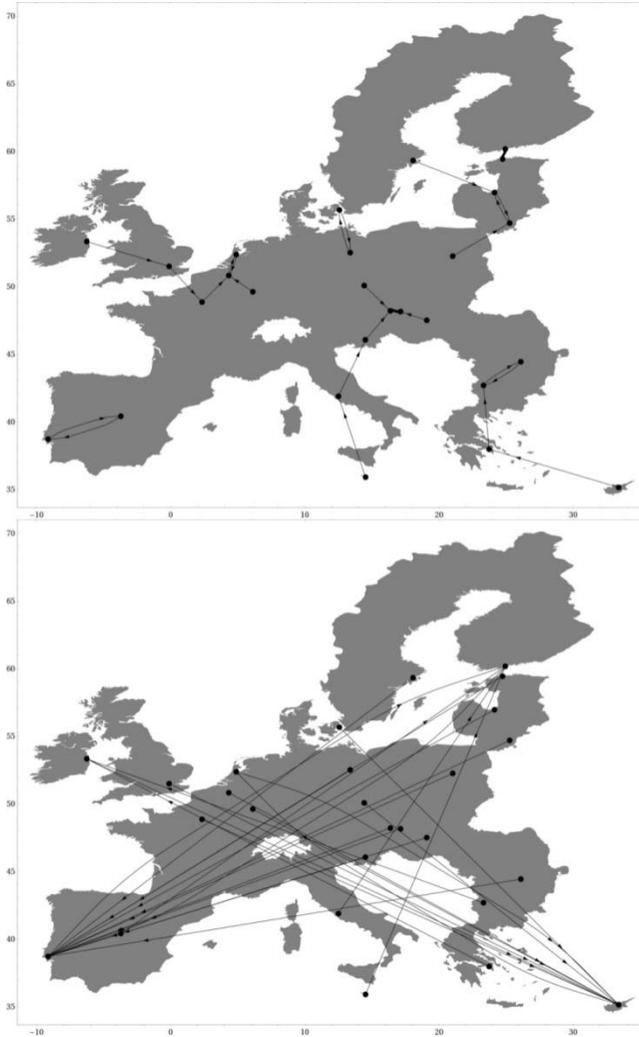


Fig. 2. Two more examples of a hypothetical EU27 trading network ($N=27$). The figure represent the maximally shrunk (top) and the maximally stretched (bottom) network, under the constraint that each vertex has at most one out-going link. The values of the filling coefficient are now $f = 0.06$ (top) and $f = 0.87$ (bottom).

The effectiveness of the encoded information can also be tested over time, by analyzing different temporal snapshots of the same network, thus highlighting the presence of statistically significant trends.

Our null models are statistical ensembles of graphs with specified properties, or *constraints*. A *graph ensemble*, \mathfrak{K} , is a collection of graphs. For our purposes, we identify \mathfrak{K} as the so-called “grand-canonical” ensemble of binary directed networks, i.e. all the networks with a given number of nodes, N , and a number of links varying from 0 to $N(N - 1)$.

We want to construct a probability measure $P(G)$, associated to each graph G of this ensemble, that allows us to realize the desired constraints (for instance, the average number of links can be set to a given value), while leaving the unconstrained properties maximally random. This is achieved by maximizing *Shannon’s entropy* (Shannon 1948)

$$S = - \sum_{G \in \mathfrak{K}} P(G) \ln P(G) \quad (3)$$

subject to the normalization condition $\sum_{G \in \mathfrak{K}} P(G) = 1$ and to the condition that a set of desired properties $\{\pi_a\}$ is realized, i.e. the expected value

$$\langle \pi_a \rangle \equiv \sum_{G \in \mathfrak{K}} P(G) \pi_a(G) \quad (4)$$

can be tuned to any desired value. The result of this constrained entropy maximization is an occurrence probability of the form

$$P(G|\vec{\theta}) = \frac{e^{-H(G,\vec{\theta})}}{Z(\vec{\theta})} \quad (5)$$

where $\vec{\theta}$ is a vector of unknown *Lagrange multipliers*, $H(G,\vec{\theta}) = \sum_a \theta_a \pi_a(G)$ is the *graph Hamiltonian* (a linear combination of the chosen constraints) and $Z(\vec{\theta}) = \sum_{G \in \mathfrak{K}} e^{-H(G,\vec{\theta})}$ is the *partition function* (Park and Newman 2004).

Given an observed graph, G^* , the Lagrange multipliers are set to the numerical values $\vec{\theta}^*$ that maximize the *log-likelihood function* defined as $L(\vec{\theta}) \equiv \ln P(G^*|\vec{\theta})$ (Garlaschelli and Loffredo 2008; Squartini and Garlaschelli 2011):

$$\left. \frac{\partial L(\vec{\theta})}{\partial \theta_a} \right|_{\vec{\theta}^*} = 0 \quad \forall a \quad (6)$$

This leads to the system of equations

$$\langle \pi_a \rangle = \pi_a(G^*) \quad \forall a. \quad (7)$$

In other words, the parameters $\vec{\theta}^*$ ensure that the expected values of the desired constraints equal the particular values observed in the real network. If inserted into $P(G|\vec{\theta})$, these parameters allow us to calculate analytically the expected value $\langle X \rangle$ of any other (unconstrained) topological property X of interest. Comparing $\langle X \rangle$ with the observed value $X(G^*)$ finally allows us to conclude whether the enforced constraints are (partially) responsible also for other unconstrained properties (Squartini and Garlaschelli 2011).

For our purposes, the above step is the key ingredient we will exploit in order to check whether a non-spatial null model (i.e. one where the chosen constraints are purely topological and independent of distances) can account for (part of) the spatial organization of a real network by filtering out spurious spatial effects and highlighting the genuine effects of distances. Note that our use of the terms “spatial”, “non-spatial” and “topological” is somewhat improper but very practical; we will give a complete clarification of our terminology at the end of the chapter. In our binary analyses, we will employ three non-spatial null models: the *Directed Random Graph* model (DRG), the *Directed Configuration Model* (DCM) and the *Reciprocated Configuration Model* (RCM) (Squartini and Garlaschelli 2011). These models are of increasing complexity, and are briefly described below.

The DRG is characterized by only one constraint: the total number of observed links, L^* . The DRG Hamiltonian is thus $H(G, \theta) = \theta L(G)$ and, for any pair of vertices i and j , the probability p_{ij} of connection is equal to

$$p_{ij} = \frac{x^*}{1 + x^*} = \frac{L^*}{N(N-1)} \quad (8)$$

where $x \equiv e^{-\theta}$ (x^* is the fitted value corresponding to θ^*). The DRG represents the simplest binary model (Erdős and Renyi 1959; Gilbert 1959).

The DCM is a more refined null model, defined by the network’s *in-degree sequence* (the vector of the in-degrees of each vertex, i.e. the numbers of incoming links defined as $k_i^{in} = \sum_{j=1}^N a_{ji}$) and *out-degree sequence* (the vector of the out-degrees of each vertex, i.e. the numbers of outgoing links defined as $k_i^{out} = \sum_{j=1}^N a_{ij}$). The DCM is one of the most

used null models in network theory and it was shown to replicate many properties of the WTW (Squartini and Garlaschelli 2011; Squartini Fagiolo and Garlaschelli 2011a). The resolution of the DCM $2N$ equations leads to a probability matrix whose generic entry has the functional form

$$p_{ij} = \frac{x_i^* y_j^*}{1 + x_i^* y_j^*} \quad (9)$$

where $x_i \equiv e^{-\theta_i^{out}}$ and $y_i \equiv e^{-\theta_i^{in}}$, θ_i^{out} and θ_i^{in} being the Lagrange multipliers coupled with the out-degree and in-degree sequences respectively (x_i^* and y_i^* denote the fitted values).

Finally, the RCM is characterized by $3N$ constraints, decomposing the in-degree and out-degree sequences into three more detailed sequences distinguishing between reciprocated (by a mutual link in the opposite direction) and non-reciprocated links. The three sequences are the following: the one of *reciprocated degrees* (the numbers of reciprocated links involving each vertex), the one of *non-reciprocated out-degrees* (the numbers of non-reciprocated out-going links from each vertex) and the one of *non-reciprocated in-degrees* (the numbers of non-reciprocated incoming links into each vertex) (Garlaschelli and Loffredo 2004; Garlaschelli and Loffredo 2006). The equations to be solved are now $3N$ and the connection probability is

$$p_{ij} = \frac{x_i^* y_j^* + z_i^* z_j^*}{1 + x_i^* y_j^* + x_j^* y_i^* + z_i^* z_j^*} \quad (10)$$

where $x_i \equiv e^{-\theta_i^{\rightarrow}}$, $y_i \equiv e^{-\theta_i^{\leftarrow}}$, $z_i \equiv e^{-\theta_i^{\leftrightarrow}}$, with θ_i^{\rightarrow} , θ_i^{\leftarrow} and $\theta_i^{\leftrightarrow}$ being the Lagrange multipliers associated with the three types of enforced node degrees (Squartini and Garlaschelli 2011).

The aforementioned three models are characterized by some kind of topological property (such as the link density, the degree sequence and the reciprocity) that is *a priori* independent of any spatial constraint. They therefore allow us to improve our definition of filling coefficient by filtering out the spurious spatial effects due to the non-spatial constraint enforced. In order to achieve this result, a comparison between the observed value of f and its expectation is needed. Consider the expected value of the filling coefficient under any of the three aforementioned null models (*NM*)

$$\langle f \rangle_{NM} = \frac{\sum_{i=1}^N \sum_{j \neq i} p_{ij}^{NM} d_{ij} - F_{\min}}{F_{\max} - F_{\min}} \quad (11)$$

where p_{ij}^{NM} is given by one of eqs. (9-11). The comparison between observation and expectation can be easily carried out by making use of the following rescaled version of the filling coefficient, that we denote as *filtered filling* (Ruzzenenti et al. 2012):

$$\varphi_{NM} \equiv \frac{f - \langle f \rangle_{NM}}{1 - \langle f \rangle_{NM}}. \quad (12)$$

The range of φ_{NM} is $[-1, 1]$. A positive value of φ_{NM} means that the considered network is “more stretched” than its expected counterpart defined by imposing a selected set of constraints on the graph ensemble. On the other hand, φ_{NM} is negative for networks which are “more shrunk” than expected. Thus, the filtered filling combines the model’s prediction and the observed information in such a way that their comparison can be carried out by simply looking at the sign of φ_{NM} . Note that the normalization in eq. (12) also allows for a comparison between networks with different topological properties (i.e. number of nodes, number of links, degree sequences, etc.), discounting for the different impact of the imposed constraints on the considered topologies. We also note that the comparison between the observed and expected values of the filling make the choice of F_{\min} and F_{\max} irrelevant, in accordance with our previous comment about the arbitrariness of the latter.

Local Measures

The filling coefficient and the filtered filling are global quantities measuring the extent to which spatial effects shape the graph as a whole. However, from our introductory remarks it is clear that a vertex-specific definition is also necessary in order to isolate potentially conflicting contributions of individual nodes. To this end, a local measure is naturally induced by the sums

$$F_i^{\text{out}} \equiv \sum_{j \neq i}^N a_{ij} d_{ij} \quad F_i^{\text{in}} \equiv \sum_{j \neq i}^N a_{ji} d_{ij}. \quad (13)$$

As before, after rescaling F_i^{out} , we can define the *local outward filling coefficient* (Ruzzenenti et al. 2012) as

$$l_i^{\text{out}} \equiv \frac{\sum_{j \neq i}^N a_{ij} d_{ij} - (F_i^{\text{out}})_{\min}}{(F_i^{\text{out}})_{\max} - (F_i^{\text{out}})_{\min}} \quad (14)$$

where the values $(F_i^{\text{out}})_{\min}$ and $(F_i^{\text{out}})_{\max}$ characterize the extreme local values for the maximally shrunk and maximally stretched configurations, in a properly-defined scenario. In analogy with the global quantities F_{\min} and F_{\max} , we choose the extreme values $(F_i^{\text{out}})_{\min}$ and $(F_i^{\text{out}})_{\max}$ as the sums of the first $\lfloor L/N \rfloor$ smallest and largest distances (now defined locally for each vertex i) respectively. This number of addenda is chosen to be consistent with the choice made at the global level: for a network with a given number L of links, the expected number of (either in-coming or out-going) connections of each node is L/N . Similarly, we can define the *local inward filling coefficient* as

$$l_i^{\text{in}} \equiv \frac{\sum_{j \neq i}^N a_{ji} d_{ij} - (F_i^{\text{in}})_{\min}}{(F_i^{\text{in}})_{\max} - (F_i^{\text{in}})_{\min}}. \quad (15)$$

Note that, due to the symmetry of the matrix of distances, $(F_i^{\text{in}})_{\max} = (F_i^{\text{out}})_{\max}$ and $(F_i^{\text{in}})_{\min} = (F_i^{\text{out}})_{\min}$.

As for the global quantity, the expected value of the local filling coefficients can be simply obtained by replacing the term a_{ij} in eqs. (14) and (15) with the probability p_{ij}^{NM} under the chosen null model. It is already very useful to compare the observed and expected values of the local filling coefficients as functions of the corresponding non-spatial properties (i.e. the out-degree or in-degree). In this case, we do not introduce any rescaled or “filtered” measure for brevity.

The effects of distances on the binary WTW

We now come to the application of the above methodology to the WTW. We analyzed the yearly binary snapshots of the network from 1948 to 2000, extracted from a comprehensive dataset (Gleditsch 2002). During this temporal interval, the number of nodes (countries) increased from $N_{1948} = 86$ to $N_{2000} = 186$, and the link density, $c_y = L/N(N-1)$, raised from $c_{1948} = 0.39$ to $c_{2000} = 0.57$. By contrast, the *average*

distance, $\mu_y = \sum_{i<j} 2d_{ij}/N(N-1)$, remained quite stable from $\mu_{1948} = 7516$ km to $\mu_{2000} = 7550$ km. This is not surprising, considering that the Earth's surface is a bounded space.

The global filtered filling φ_{NM} , calculated under the three null models, is plotted as a function of time in **Fig. 3**. In the period under consideration, all null models always yield negative values of φ_{NM} . This means that the WTW is a systematically “shrunk” network, confirming the naïve expectation that geographic distances have a suppressing effect on trade: the farthest the countries, the lesser the probability to observe a trade exchange between them (remember that we are carrying out a binary analysis for the moment). However, the small measured values ($-0.250 \leq \varphi_{NM} \leq 0$) seem also to suggest that the role played by distances is quite weak, a result that appears to contrast classical economic arguments (Linders Martijn and Van Oort 2008).

While the three models qualitatively agree in classifying the WTW as spatially shrunk, we observe important quantitative differences both among models and over time. The temporal trends obtained under the RCM and the DCM are practically identical, but (from 1960 onwards) they are almost inverted with respect to the trend obtained under the DRG.

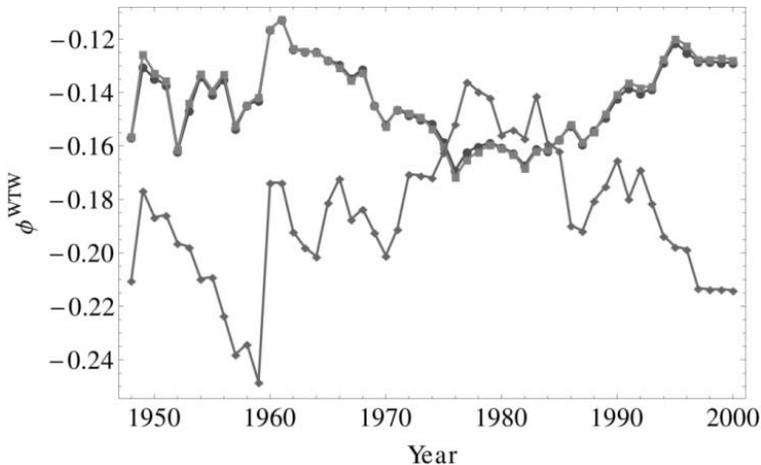


Fig. 3. The filtered filling coefficient φ_{NM} of the binary WTW from year 1948 to 2000, under the three null models considered: DRG (diamonds), DCM (circles) and RCM (squares).

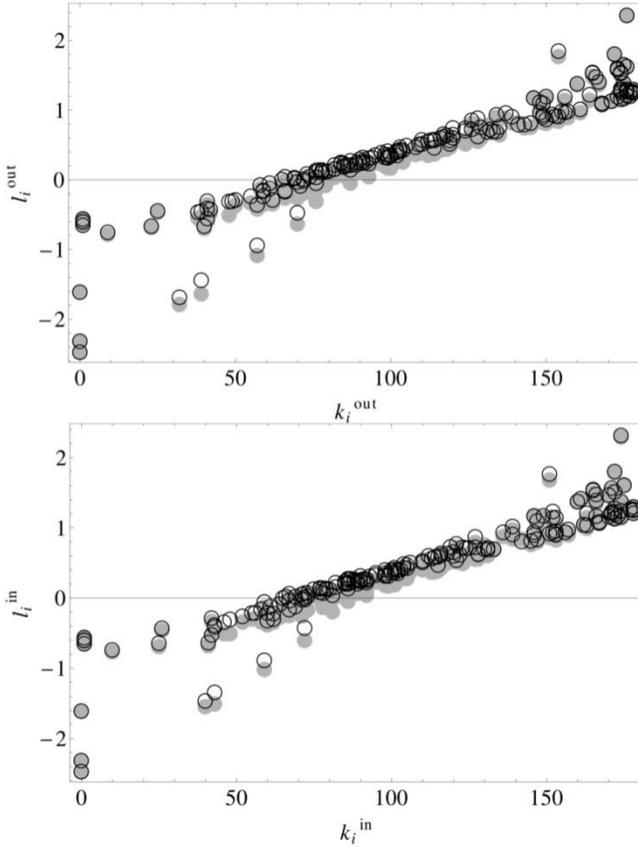


Fig. 4. Local outward filling, defined in eq.(14), versus out-degree (top panel) and local inward filling, defined in eq.(15), versus in-degree (bottom panel). The empty circles represent the observed values, while the filled circles represent the expected values predicted by the DCM.

The first finding means that the introduction of reciprocity as an additional constraint is not really necessary in order to filter out the local non-spatial effects, which seem to be already effectively discounted by the in- and out-degree sequences alone.

A naive explanation might be the high symmetry of the WTW, i.e. the high number of reciprocated interactions between world countries (Ruzzenenti Garlaschelli and Basosi 2010; Garlaschelli and Loffredo

2004). This high reciprocal structure, which reduces the WTW almost to an undirected network, makes the information carried by the reciprocity irrelevant. However, as we show later, this interpretation is incorrect. A statistically appropriate procedure to quantify and rank the effectiveness of different models in explaining the observed network structure is presented in the third section of this chapter. Its application reveals that the reciprocity is a key and irreducible structural property of the WTW (Picciolo et al. 2012).

The second finding, i.e. the almost inverted trend of the DRG with respect to the other two models, is a result of the intrinsic difference between the homogeneity of the DRG (which controls only for the overall density of trade) and the heterogeneity of the other models (which control for country-specific properties). The continuous appearance of unrealized long-distance connections overcompensates the establishment of a few new ones, and the overall result is an effective shrinking of the network. At this point, it is worth mentioning that the topology of the real WTW is very different from that of the DRG, while it is instead accurately reproduced by the DCM and especially the RCM (Squartini Fagiolo and Garlaschelli 2011a; Squartini and Garlaschelli 2013). This means that the non-spatial effects filtered out by the DRG do not represent key structural properties shaping the real WTW.

By contrast, the DCM and RCM filter out the most informative properties, i.e. the ones that are sufficient in order to reproduce the observed topology of the WTW. The use of the DCM and RCM should therefore be strongly preferred to that of the DRG when trying to disentangle spatial and non-spatial effects in the WTW. The empirical inverted trends shown above warn us about the opposite interpretations that can arise from a misuse of homogeneous network benchmarks.

Focusing on the trend obtained under the heterogeneous models, we find that the two periods known in the economic literature as the first and second “waves” of globalization (De Benedictis and Helg 2002; Crafts 2004) turn out to correspond to two opposite phenomena at a topological level. During the “first wave”, i.e. the period starting around 1960 during which many former colonies became independent states, the topology of the WTW actually “shrank”.

This result is apparently a paradox, since it is known that the new independent states (which gradually appear as new nodes in the network) kept a strong trade relationship with their former colonizers, thus originating new long-distance links and (in principle) “stretching out” the WTW. However, one must also note that the appearance of the new nodes, while accompanied by new long-distance links, is also accompanied by

many new *missing* long-distance links: two new (and generally small) independent states located at opposite locations on the globe typically do not trade with each other.

By contrast, during the “second wave” of globalization corresponding to the fall of the east-west division in Europe and the disintegration of the Soviet Union, the WTW stretched out topologically, as indicated by the rise of the trend between the late Eighties and the mid Nineties. Since the trade relationships linking the formerly Soviet states are short distance the overall stretching of the WTW must be the result of the establishment of additional long-distance connections. In other words, unlike the previous phase, the new states are now really internationally integrated, at least at a topological level.

We now turn to a local analysis of spatial effects. The local spatial quantities defined in eqs. (14) and (15) are plotted as a function of the corresponding non-spatial properties in **Fig. 4**. The top panel shows the local outward filling plotted versus the out-degree, while the bottom panel shows the local inward filling plotted versus the in-degree. We show the results for the year 2000 only, but similar results are observed for all the considered years. The expected values under the DCM are also plotted; we do not show the expected values according to the RCM because they overlap to the DCM ones. We find a strong nonlinear correlation between node degrees and local filling values (both outward and inward). For countries with very small and very large degrees, the accordance with the null model is almost perfect, while the largest discrepancy is observed for countries with intermediate values of the degree. Our explanation of this effect is the following. Countries with degree (almost) equal to the maximum value are necessarily connected with (almost) every other country, both in the real network and in the null model (because the latter preserves the number of links of each node). This generates the accordance with the null model for large-degree countries, and also for small-degree ones: the latter countries are in turn necessarily connected with the “hubs”, irrespective of distances. Only countries in the intermediate range of connectivity can have a large degree of freedom. The figure shows that these countries have systematically a stronger than predicted tendency to trade with geographically closer countries.

The global spatial effects discussed above, encapsulated in a negative value of the filtered filling, come only from these intermediate-degree countries, and they are therefore not representative of the behavior of all nodes.

Spatial effects in weighted networks

The concepts introduced in the previous section can be generalized to the weighted case (Ruzzenenti et al. 2012).

A weighted graph can be unambiguously defined by an adjacency matrix, W , whose generic entry, w_{ij} , represents the *intensity* of the link from node i to node j (we assume again that self-loops are absent, i.e. $w_{ii} = 0$). In this section, we first define the weighted counterparts of the quantities we have already introduced (this also include a definition of weighted null models). Later on we present the corresponding application to the analysis of the WTW as a weighted network.

Weighted definitions

By looking at eq. (1), we can define the weighted analogous of F as

$$F^w = \sum_{i=1}^N \sum_{j \neq i} w_{ij} d_{ij} \quad (16)$$

Similarly, the *weighted filling coefficient* can be written as

$$f^w = \frac{\sum_{i=1}^N \sum_{j \neq i} w_{ij} d_{ij} - F_{\min}^w}{F_{\max}^w - F_{\min}^w}. \quad (17)$$

Also in the weighted case, the two extreme values of F^w can be chosen in an arbitrary way. For instance, if we fix the total weight $W = \sum_{i=1}^N \sum_{j \neq i} w_{ij}$, F^w reaches its lowest and highest value when W is placed between the two nearest and farthest vertices respectively, i.e. $F_{\min}^w = W d_1^\dagger$ and $F_{\max}^w = W d_1^l$ (Ruzzenenti et al. 2012).

As for binary networks, we can introduce null models in order to have a benchmark filtering out non-spatial effects. The *Weighted Random Graph* model (WRG) is the analogous of the DRG for binary networks. The only constraint we impose is the total weight, W , and the Hamiltonian is $H(G) = \theta W(G)$.

The expected weight of the link from node i to node j is

$$\langle w_{ij} \rangle = \frac{y^*}{1 - y^*} = \frac{W}{N(N-1)} \quad (18)$$

where now $y \equiv e^{-\theta}$ (y^* is the fitted value corresponding to θ^*). By imposing this only constraint, we are exclusively making use of the average intensity of the links (Garlaschelli 2009).

The second weighted null model we consider is the *Weighted Configuration Model* (WCM), where the constraints are the *in-strength* and *out-strength sequences*, defined by the $2N$ values of the *in-strength*, $s_i^{\text{in}} = \sum_{j=1}^N w_{ji}$, and the *out-strength*, $s_i^{\text{out}} = \sum_{j=1}^N w_{ij}$, of vertices. The expected link is now

$$\langle w_{ij} \rangle = \frac{x_i^* y_j^*}{1 - x_i^* y_j^*} \quad (19)$$

where $x_i \equiv e^{-\theta_i^{\text{out}}}$ and $y_i \equiv e^{-\theta_i^{\text{in}}}$ (x_i^* and y_i^* indicate the fitted values realizing the observed strength sequences) (Squartini and Garlaschelli 2011).

Even if a more comprehensive list of null models for weighted networks has been defined recently, we will consider only the WRG and WCM for brevity (Squartini et al. 2013; Mastrandrea et al., 2013).

As in the binary case, these models allow us to obtain the expected value of the weighted filling coefficient by simply substituting in eq. (17) the observed link weight, w_{ij} , with the expected one, $\langle w_{ij} \rangle_{NM}$, calculated using either eq. (18) or eq. (19). The observed and expected values can be combined in the following definition of *weighted filtered filling*:

$$\varphi_{NM}^w \equiv \frac{f^w - \langle f^w \rangle_{NM}}{1 - \langle f^w \rangle_{NM}} \quad (20)$$

which again ranges between -1 and $+1$. A positive (negative) value of φ_{NM}^w means that distances have a stretching (shrinking) effect on the link weights of the observed weighted network (Ruzzenenti et al. 2012).

A final extension concerns the local structure. The sums

$$F_i^{w\text{out}} \equiv \sum_{j \neq i}^N w_{ij} d_{ij} \quad F_i^{w\text{in}} \equiv \sum_{j \neq i}^N w_{ji} d_{ij} \quad (21)$$

lead us to the following definitions of *local outward weighted filling* (Ruzzenenti et al. 2012)

$$l_i^{w\text{out}} \equiv \frac{\sum_{j \neq i}^N w_{ij} d_{ij} - (F_i^{w\text{out}})_{\min}}{(F_i^{w\text{out}})_{\max} - (F_i^{w\text{out}})_{\min}} \quad (22)$$

and *local inward weighted filling*

$$l_i^{w\text{in}} \equiv \frac{\sum_{j \neq i}^N w_{ji} d_{ij} - (F_i^{w\text{in}})_{\min}}{(F_i^{w\text{in}})_{\max} - (F_i^{w\text{in}})_{\min}} \quad (23)$$

where the minimum and maximum values for $F_i^{w\text{out}}$ and $F_i^{w\text{in}}$ characterize the maximally stretched and shrunk possibilities for vertex i , in a properly chosen weighted scenario. In analogy with the choice made for the global quantity, we choose a scenario where the total weight W is fixed. The resulting expected in-strength and out-strength of every vertex have the same value W/N . In a straightforward approach, our choice for the extreme values of $F_i^{w\text{out}}$ and $F_i^{w\text{in}}$ is such that vertex i concentrates all its out-strength in a single outgoing link of weight W/N directed to the spatially closest vertex, and all its in-strength in a single incoming link of weight W/N coming from the same vertex. Note that this implies $(F_i^{w\text{in}})_{\max} = (F_i^{w\text{out}})_{\max}$ and $(F_i^{w\text{in}})_{\min} = (F_i^{w\text{out}})_{\min}$. As above, the expected local outward filling is simply obtained by replacing the terms w_{ij} in eqs. (22) and (23) with the expectations $\langle w_{ij} \rangle$ under the chosen null model.

The effects of distances in the weighted WTW

We can now perform a new analysis of the WTW, by considering its weighted structure rather than its topology. Our weighted analysis spans again the years from 1948 to 2000.

As shown in **Fig. 5**, the (small) negative values of the global filtered filling confirm that the WTW is a (weakly) shrunk network. However, the temporal trends are very different from the corresponding binary ones.

Surprisingly, according to the WCM, the strongest spatial stretching occurred during the Fifties, while during the first wave of globalization the trend remained approximately constant. The second wave of globalization corresponds instead to a decreasing trend, now signaling an unexpected spatial shrinking of the network.

The WRG is instead more in line with the DCM, and identifies a shrinking during the first wave and a sudden stretching during the second wave. Considering together the binary and weighted results, it appears

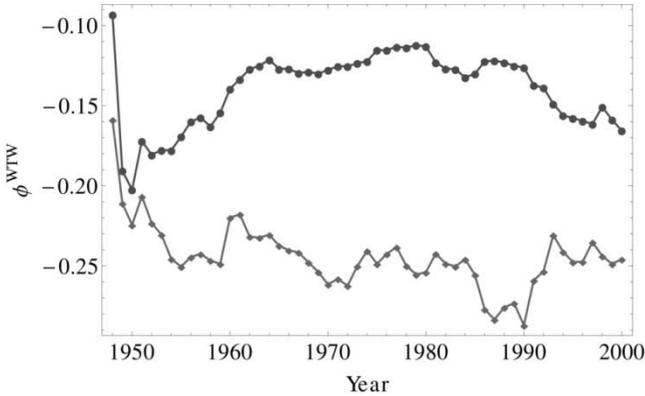


Fig. 5. The filtered weighted filling for the WTW from year 1948 to 2000, under the two null models WRG (diamonds) and WCM (circles).

that two tendencies coexist. First, the WTW topology has become more and more stretched during the last decade of the sample, with distances opposing less and less resistance. Second, the intensity of trade exchanges has risen more between countries that are geographically closer, with distances opposing more and more resistance.

In other words, it appears that during the last wave of globalization the WTW has, from an “extensive” point of view, tended to stretch out in its embedding space by effectively preferring long-distance connections, and, from an “intensive” point of view, tended to shrink in by strengthening the existing links between close neighbors.

However, the above results must be interpreted with particular care, since (unlike the DCM) both the WRG and the WCM are known to be very poor models of the WTW (Squartini Fagiolo and Garlaschelli 2011b). We therefore warn the reader that the WCM does not filter out the weighted, non-spatial patterns as satisfactorily as the DCM does. In order to reproduce the weighted structure of the WTW, a more refined model combining binary and weighted constraints is needed (Mastrandrea et al., 2013; Squartini and Garlaschelli 2013). Thus, even if from an economic point of view the WCM might appear more satisfactory than the DCM, because it controls for the total imports and exports of countries, it turns out to be uninformative about other properties of the network.

Counter-intuitively, the number of exporters and importers (which defines the DCM), turns out to be a much more informative property. We will comment again on this point when discussing Gravity Models at the end of the chapter.

Keeping the above warning in mind, we finally consider the local spatial effects in the weighted WTW. The top panel of **Fig. 6** shows both the observed and expected local weighted outward filling, plotted versus the out-strength sequence, while the bottom panel shows both the observed and expected local weighted inward filling, plotted versus the in-strength sequence. We only show the results for the year 2000, but similar results are observed for all the considered time period.

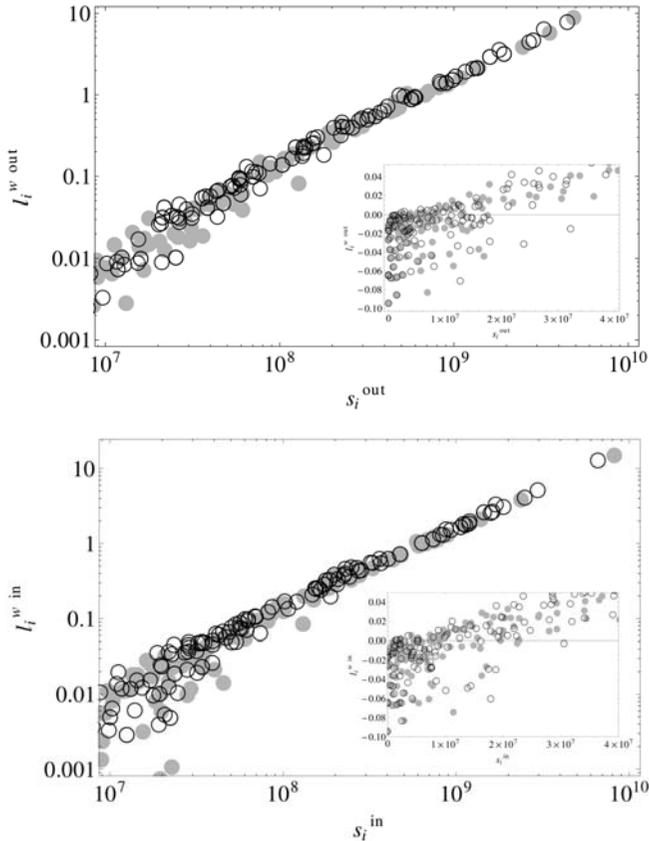


Fig. 6. Local outward weighted filling versus out-strength (top panel) and local inward weighted filling versus in-strength (bottom panel). The empty circles represent the observed values, while the filled circles represent the expected values predicted by the WCM.