



A mathematical model for networks with structures in the mesoscale

Regino Criado, Julio Flores, Alejandro Jose Gacia del Amo, Jesus Gómez, Miguel Romance

► To cite this version:

Regino Criado, Julio Flores, Alejandro Jose Gacia del Amo, Jesus Gómez, Miguel Romance. A mathematical model for networks with structures in the mesoscale. *International Journal of Computer Mathematics*, 2011, pp.1. 10.1080/00207160.2011.577212 . hal-00710059

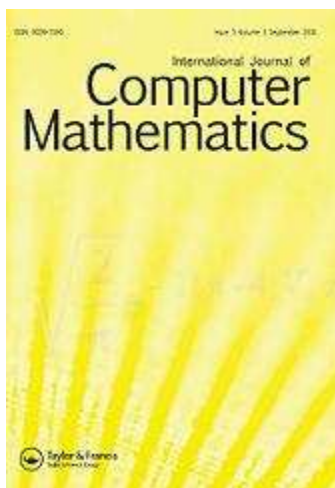
HAL Id: hal-00710059

<https://hal.science/hal-00710059>

Submitted on 20 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A mathematical model for networks with structures in the mesoscale

Journal:	<i>International Journal of Computer Mathematics</i>
Manuscript ID:	GCOM-2010-0956-A.R2
Manuscript Type:	Original Article
Date Submitted by the Author:	03-Mar-2011
Complete List of Authors:	Criado, Regino; Universidad Rey Juan Carlos, Matematica Aplicada Flores, Julio; Universidad Rey Juan Carlos, Matematica Aplicada Gacia del Amo, Alejandro; Universidad Rey Juan Carlos, Matematica Aplicada Gómez, Jesus; Universidad Rey Juan Carlos, Matematica Aplicada Romance, Miguel; Universidad Rey Juan Carlos, Matematica Aplicada
Keywords:	multilevel networks, clustering, efficiency, mesoscale, random models

SCHOLARONE™
Manuscripts

A mathematical model for networks with structures in the mesoscale

Regino Criado ^{*}, Julio Flores [†], Alejandro García del Amo [‡]
Jesús Gómez-Gardeñes [§] and Miguel Romance [¶]

Department of Applied Mathematics
Rey Juan Carlos University
28933 Móstoles (Madrid), Spain

Abstract

The new concept of multilevel network is introduced in order to embody some topological properties of complex systems with structures in the mesoscale which are not completely captured by the classical models. This new model, which generalizes the hyper-network and hyper-structure models, fits perfectly with several real-life complex systems, including social and public transportation networks. We present an analysis of the structural properties of the multilevel network, including the clustering and the metric structures. Some analytical relationships amongst the efficiency and clustering coefficient of this new model and the corresponding parameters of the underlying network are obtained. Finally some random models for multilevel networks are given to illustrate how different multilevel structures can produce similar underlying networks and therefore that the mesoscale structure should be taken into account in many applications.

1 Introduction

During the last years the scientific community has shown that many relevant properties of communication systems, social networks and other biological and technological systems may be described in terms of complex network properties, including structural and dynamical properties and the interplay between them (see, for example, [1], [2], [5],[7], [8], [14] or [19] and the references therein). The emergence of

^{*}e-mail: regino.criado@urjc.es
[†]e-mail: julio.flores@urjc.es
[‡]e-mail: alejandro.garciadelamo@urjc.es
[§]e-mail: jesus.gomez.gardenes@urjc.es
[¶]e-mail: miguel.romance@urjc.es

these appealing results has forged the Complex Networks Analysis as an attractive and multidisciplinary branch of research that includes topics of sociology (social networks), biological sciences (including metabolic and protein networks, genetic regulatory network and many others), neuro-sciences (neural interaction networks, cortical networks) engineering (phone call networks, computers in telecommunication networks, Internet, the World Wide Web), theoretical physics and applied mathematics.

The working objects of the Complex Networks Analysis are graphs (which are called *complex networks*) which occasionally prove unable to capture the details present in some real-life problems. One example of this situation can be found in sociology. Social networks analysis is used in the social and behavioral sciences, as well as in economics, marketing, and industrial engineering ([20]), but some questions related to the structure of social networks have been not properly understood. From the fact that a social network can be understood as a set of people or groups of people with some pattern of contacts or interactions between them ([18],[20]) - think of the Facebook network - it might seem that all the connections or social relationships between the members of that network take place at the same level. But the real situation is far from this. The real relationships amongst the members of a social network take place (mostly) inside of different groups (levels) and therefore they cannot be properly modeled if only the natural local-scale point of view used in classic complex network models is taken into account. This phenomenon is related to the fact that several real-life complex systems have a community structure [11] that takes place in a scale that lies between the local and the global one and that gives information about the structure-functionality relationship. The problem of finding and analyzing the community structure of a complex network is one of the *hot spots* of the complex network analysis, since these meso-scale structures reveal valuable information about roles of groups of nodes [15].

The main goal of this paper is to present the concept of multilevel network as a new model useful to analyze with detail complex systems that have some structures in the mesoscale level. As we pointed out before, this kind of networks are present in many real-life situations, not only in social networks, but also in communication networks and all the other examples that exhibit a community structure. The introduction of a sharp mathematical model that fits this new structures may prove crucial to properly analyze the dynamics that take place in these complex systems which are far from being completely understood. An example of these phenomena is the dissemination of culture in social networks by using the Axelrod Model that has been recently studied by using meso-scale networks [12]. Yet the study of other meso-scaled complex networks and their dynamics is one of the scientific challenges of the nowadays complex network analysis. There are other mathematical models that consider non-local structures such as the hypergraphs (or hypernetworks) [4] or the hyper-structures [9] but they are not able to combine the local scale with the global and the mesoscale structures of the system. For example, if we want to model how a rumor is spread within a social network, it is necessary to have in mind

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

not only that different groups are linked only through some of their members but also that two people who know the same person don't have necessarily to know each other, and, in fact, these two people may belong to different groups or levels. If we try to model this situation with hyper-networks, then we will only be taking into account the social groups and not the actual relationships between their members; in contrast if we use the hyper-structure model, then we won't possibly know to which social group each contact between two nodes belongs.

After multilevel networks are defined, we present a structural analysis of these objects in terms of the clustering coefficient and the metric structure which must be properly defined in this setting. We will compare these parameters with the corresponding ones for the underlying projection networks and we will present the analytical relationships between them. Despite the fact that there are sharp connections between the structural properties of the multilevel network and the properties of their projection and layer (or slice) networks, we will present some examples that exhibit the deep differences between the classic complex network approach and this new point of view.

Finally, in section 3, we will give several growing models to produce multilevel random networks, inspired by the Barabási-Albert preferential attachment model of complex networks [3]. These methods are inspired by the bipartite networks models introduced by Ramasco *et al.* [17] and by M. Peltomäki and M. Alava [16], but in this case we have to take into account the meso-scale nature of the multilevel structure. The use of growing random models is natural since many of the examples in real life that can be modeled by multilevel networks, such as social or transportation networks, are dynamic models whose structure **grows** in time by addition of new nodes, links and layers. By using these models we perform several random testings that show that we can get very different multilevel networks (i.e. with very different meso-scale structure) which still have similar local and global structures when considered as classic complex network. For example, we exhibit a multilevel network that at the meso-scale is like an Erdős-Rényi random network, but it has a scale-free structure if we consider it as a complex network. These random testings reinforce the idea that in many cases we must take into account the meso-scale structure in order to get deep understanding of the structure and dynamics of many complex systems.

2 Multilevel Networks: A Structural Analysis

As we pointed out in the previous section, a sort of naive approach to complex systems with structures in the mesoscale could suggest that hyper-networks and hyper-structures fit perfectly to these real-life systems. The key-point that makes these mathematical model not to be the best solution for systems with structures in the mesoscale has to do with the fact that both hyper-networks and hyper-structures are node-based models, while many real systems combine a node-based point of

view with a link-based perspective. For example, if we have a look again at a social network with structures in the mesoscale (such as several social groups within its structure), when we consider a relationship between two members of one social group(or several), we have to take into account not only the social groups that hold the members, but also the social **groups** that hold the relationship itself, i.e., if, for example, there is a relationship between two people that share the same working group and the same sport group, we have to highlight if the relationship is due to sharing the same group of work or it has sport nature. This fact is not particular of social networks with structures in the mesoscale and a similar situation occurs, for example, in public transportation systems, where a link between two stations belonging to several transport lines can occur as a part of different lines.

In order to avoid this node-based nature of hyper-networks and hyper-structures, we propose to introduce the following concept that combines the node-based with the link-based perspective:

Definition 2.1 Let $G = (X, E)$ be a network. A *multilevel network* is a triple $M = (X, E, \mathcal{S})$, where $\mathcal{S} = \{S_1, \dots, S_p\}$ is a family of subgraphs $S_q = (X_q, E_q)$ of G such that

$$(i) \quad X = \bigcup_{j=1}^p X_j,$$

$$(ii) \quad E = \bigcup_{j=1}^p E_j.$$

The network G is the *projection network* of M and each subgraph $S_j \in \mathcal{S}$ is called a *slice* of the multilevel network M .

This mathematical model perfectly suits social systems (as well as other complex systems) with structures in the mesoscale, since each social group can be understood as a *slice graph* in a *multilevel network*; thus we are simultaneously taking into account the nature of the links (i.e. relationships) and the nodes involved.

It is easy to check that this new mathematical object extends both the classic complex network model and also the hyper-network model [4]. Let us point it out very briefly. On the one hand, if $G = (X, E)$ is a network, then it can be easily seen as a multilevel network by considering $M = (X, E, \mathcal{S})$, where $\mathcal{S} = \{G\}$.

On the other hand if $\mathcal{H} = (X, H)$ is an hyper-network (i.e. X is a non-empty set of nodes and $H = \{H_1, \dots, H_p\}$ is a family of non-empty subsets of X , each of them called and *hyper-link* of $\mathcal{H}$), then it can be seen as the multilevel set $M = (X, E_{\mathcal{H}}, \mathcal{S}_{\mathcal{H}})$, given by

$$\mathcal{S}_{\mathcal{H}} = \{K_{H_1}, \dots, K_{H_p}\},$$

where $K_{H_j} = (H_j, E_j)$ is the complete network obtained by linking every pair of

nodes of H_j and

$$E_{\mathcal{H}} = \bigcup_{r=1}^p E_r.$$

By using similar arguments we can show that every hyper-structure [9] can be understood as a particular multilevel network, by considering one slice network for each hyper-link in the hyper-structure and each slice graph being a set of isolated nodes.

Once we have introduced this new and novel mathematical object, we have to give suitable structural parameters to analyze it. We can give natural extensions of many of the usual tools of the complex networks' analysis, such as the clustering coefficient, an adjacency matrix/tensor, a natural network representation as a tripartite network or a geodesic structure, among many others.

2.1 Clustering of multilevel networks

In this subsection we extend the graph clustering coefficient introduced by Watts and Strogatz in [21] to multilevel networks and we establish some relationships between the clustering coefficient of a multilevel network, the clustering coefficient of its slices and the clustering coefficient of its projection network. Recall that given a network $G = (X, E)$ the clustering coefficient of a given node i is defined as

$$c_G(i) = \frac{\text{number of links between the neighbors of } i}{\text{largest possible number of links between the neighbors of } i}.$$

We take $a/0$ to be 0 for every number a . Thus, if we think of three people i, j and k with mutual relations between i and j as well as between i and k , the clustering coefficient of i is supposed to represent the likeliness that j and k are also related. The clustering coefficient of G is usually defined as the average of the clustering coefficients of all nodes.

In order to give a definition of the clustering coefficient of a node in a multilevel network $M = (X, E, \mathcal{S})$, where $\mathcal{S} = \{S_1, \dots, S_p\}$ and $S_q = (X_q, E_q)$ for every $1 \leq q \leq p$, we need to introduce some notation. For every node $i \in X$ call $\mathcal{N}(i)$ the set of all neighbors of the node i in the projection graph G . For every $q = 1, \dots, p$ we will denote $\mathcal{N}_q(i) = \mathcal{N}(i) \cap X_q$ and $\overline{G}_q(i)$ the subgraph of the slice S_q generated by $\mathcal{N}_q(i)$, i.e. $\overline{G}_q(i) = (\mathcal{N}_q(i), \overline{E}_q(i))$, where

$$\overline{E}_q(i) = \{\{k, j\} \in E_q \mid k, j \in \mathcal{N}_q(i)\}.$$

Similarly we will define $\overline{G}(i)$ as the subgraph of the projection network G generated by $\mathcal{N}(i)$, i.e. $\overline{G}(i) = (\mathcal{N}(i), \overline{E}(i))$, where

$$\overline{E}(i) = \{\{k, j\} \in E \mid k, j \in \mathcal{N}(i)\}.$$

In addition the complete graph generated by $\mathcal{N}_q(i)$ will be denoted as usual by $K_{\mathcal{N}_q(i)}$. Note that the largest possible number of links between the nodes of $\mathcal{N}_q(i)$ is precisely the number of links in $K_{\mathcal{N}_q(i)}$, i.e.

$$|\overline{E}_q(i)| \leq \frac{|\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1)}{2},$$

where $|\mathcal{N}_q(i)|$ stands as above for the cardinality of $\mathcal{N}_q(i)$. With the previous notation we give the following

Definition 2.2 Let $M = (X, E, \mathcal{S})$ be a multilevel network. The *clustering coefficient* of a given node i in M is defined as

$$\mathbf{C}_M(i) = \frac{2 \sum_{q=1}^p |\overline{E}_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1)}.$$

The *clustering coefficient* of M is the average of all $\mathbf{C}_M(i)$ over the set of nodes.

Notice that the clustering coefficient might have been defined differently. For instance we might have considered the clustering coefficient of each S_q , say $c_q(i)$, and then take the average over the set $\{S_q\}_{q=1}^p$. However we have opted for a definition that embodies the following idea: it is possible for a given node i to have two neighbors k, j with $\{i, k\} \in E_q$, $\{i, j\} \in E_{q'}$ with $q \neq q'$ and such that there is a link $\{j, k\} \in E_{q''}$ with $q'' \neq q, q'$. This is a natural situation when we think of social networks; indeed, one person i knows j from the aerobic class, i also knows k from her reading club while j and k know each other from the supermarket. This sort of situation produces more links in the subgraphs $\overline{G}_q(i)$ defined above than those already present in the slice S_q . Taking the alternative definition based on the clustering coefficients of the slices would not help to discriminate such situations; in contrast the definition proposed does. The following example should clarify this situation and justifies the restrictions of theorem 2.4.

Example 2.3 If we consider the multilevel network $M = (X, E, \mathcal{S})$ where $X = \{1, 2, 3, 4\}$, $\mathcal{S} = \{S_1, S_2, S_3\}$ and G is as in the next figure, it is easy to check that $c_G(i) = 1$ for all $i \in X$ but however $c_{S_q}(i) = 0$ for all $i \in X$ and $q = 1, 2, 3$.

As we have seen in the last example, the clustering coefficient of the projected graph G and the clustering coefficient of each slice of M may be very different. However we are interested in establishing relations between the clustering coefficient of the multilevel network, $\mathbf{C}_M(i)$, and the clustering coefficient in the projected network, $c_G(i)$, if possible. That is the target of the next theorem:

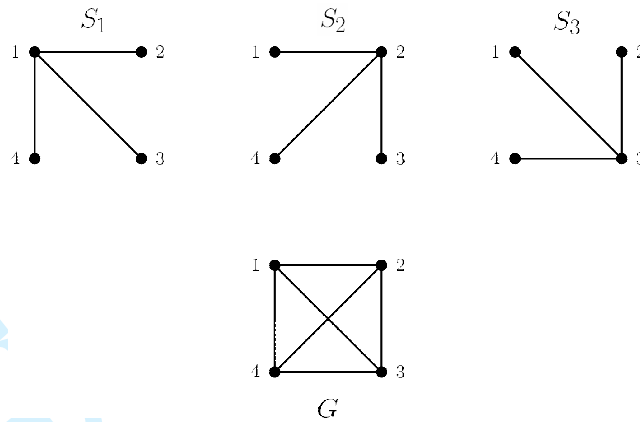


Figure 1: An example of the extreme difference between clustering coefficients

Theorem 2.4 Let $M = (X, E, \mathcal{S})$ be a multilevel network and $i \in X$. If we consider $A(i) = \{q \in \{1, \dots, p\} \mid |\mathcal{N}_q(i)| < 2\}$ and we denote $\theta(i) = |A(i)|$, then

$$\frac{1}{p - \theta(i)} c_G(i) \leq \mathbf{C}_M(i) \leq p c_G(i) \left(1 + (p - 1) \left(4 + \frac{\theta(i)}{p - \theta(i)} \right) \right). \quad (2.1)$$

Furthermore, if M is homogeneous (i.e. $X_q = X$ for every $S_q = (X_q, E_q) \in \mathcal{S}$), then

$$\frac{1}{p - \theta(i)} c_G(i) \leq \mathbf{C}_M(i) \leq c_G(i). \quad (2.2)$$

Proof:

Let us start with the left-hand inequality in (2.1). Using that $\mathcal{N}_q(i) \subseteq \mathcal{N}(i)$ for every $1 \leq q \leq p$ and $\overline{E}(i) = \cup_{q=1}^p \overline{E}_q(i)$, we get that

$$\begin{aligned} c_G(i) &= \frac{2|\overline{E}(i)|}{|\mathcal{N}(i)|(|\mathcal{N}(i)| - 1)} = \frac{\sum_{q=1}^p |\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1)}{|\mathcal{N}(i)|(|\mathcal{N}(i)| - 1)} \cdot \frac{2|\overline{E}(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1)} \\ &\leq (p - \theta(i)) \frac{2|\overline{E}(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1)} \leq (p - \theta(i)) \frac{2 \sum_{q=1}^p |\overline{E}_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1)} \\ &= (p - \theta(i)) \mathbf{C}_M(i), \end{aligned}$$

since

$$\sum_{q=1}^p |\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1) = \sum_{q \notin A(i)} |\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1) \leq (p - \theta(i))(|\mathcal{N}(i)|(|\mathcal{N}(i)| - 1)).$$

In order to prove the remaining inequality notice that the number of links in the complete graph built up on the set of nodes $\mathcal{N}(i)$ never exceeds the sum over

$q = 1, \dots, p$ of the numbers of links of the complete graphs built up on the sets of nodes $\mathcal{N}_q(i)$ plus all the links possibly obtained by joining nodes of $\mathcal{N}_k(i)$ to nodes of $\mathcal{N}_j(i)$ where k and j are chosen among all possible pairs (k, j) with $1 \leq k < j \leq p$, i.e.

$$\frac{|\mathcal{N}(i)|(|\mathcal{N}(i)| - 1)}{2} \leq \frac{1}{2} \sum_{q=1}^p |\mathcal{N}_q(i)|(|\mathcal{N}_q(i)| - 1) + \sum_{k < j} |\mathcal{N}_k(i)| |\mathcal{N}_j(i)|, \quad (2.3)$$

where the last sum has $\binom{p}{2}$ terms. Notice also that we can always assume, by rearranging if necessary, that $|\mathcal{N}_q(i)| \leq |\mathcal{N}_{q+1}(i)|$ for $1 \leq q \leq p-1$. In addition to this, observe finally that

$$\begin{aligned} \sum_{k < j} |\mathcal{N}_j(i)| (|\mathcal{N}_k(i)| - 1) &\leq \sum_{q=1}^p [(|\mathcal{N}_q(i)| - 1) |\mathcal{N}_q(i)| (q-1)] \\ &\leq (p-1) \sum_{q=1}^p (|\mathcal{N}_q(i)| - 1) |\mathcal{N}_q(i)|. \end{aligned} \quad (2.4)$$

Thus, since $\overline{E}_q(i) \subseteq \overline{E}(i)$ for every $1 \leq q \leq p$ and by using (2.3)

$$\begin{aligned} \mathbf{C}_M(i) &= \frac{2 \sum_{q=1}^p |\overline{E}_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \leq \frac{2p |\overline{E}(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \\ &= p c_G(i) \frac{|\mathcal{N}(i)| (|\mathcal{N}(i)| - 1)}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \\ &= p c_G(i) \left(1 + \frac{2 \sum_{k < j} |\mathcal{N}_k(i)| |\mathcal{N}_j(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \right). \end{aligned} \quad (2.5)$$

Notice that if $1 \leq k < j \leq p$, then

$$|\mathcal{N}_k(i)| |\mathcal{N}_j(i)| = (|\mathcal{N}_k(i)| - 1 + 1) |\mathcal{N}_j(i)| \leq (|\mathcal{N}_k(i)| - 1) |\mathcal{N}_j(i)| + |\mathcal{N}_j(i)|.$$

Now if we combine the last formula with (2.5), we get that

$$\begin{aligned} \mathbf{C}_M(i) &\leq p c_G(i) \left(1 + \frac{2 \sum_{k < j} |\mathcal{N}_k(i)| |\mathcal{N}_j(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \right) \\ &\leq p c_G(i) \left(1 + 2(p-1) + 2 \frac{\sum_{q=1}^p (q-1) |\mathcal{N}_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \right) \\ &\leq p c_G(i) \left(1 + 2(p-1) + 2 \frac{(p-1) \sum_{q=1}^p |\mathcal{N}_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \right). \end{aligned}$$

It is easy to check that if $q \in A(i)$, then $|\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1) = 0$, which makes that

$$\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1) = \sum_{q \notin A(i)} |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1) \geq \sum_{q \notin A(i)} |\mathcal{N}_q(i)|.$$

Hence

$$\begin{aligned} \frac{\sum_{q=1}^p |\mathcal{N}_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} &\leq \frac{\sum_{q \notin A(i)} |\mathcal{N}_q(i)| + \sum_{q \in A(i)} |\mathcal{N}_q(i)|}{\sum_{q \notin A(i)} |\mathcal{N}_q(i)|} \\ &\leq 1 + \frac{\theta(i)}{2(p - \theta(i))}, \end{aligned}$$

since if $q \in A(i)$, then $|\mathcal{N}_q(i)| \leq 1$, while if $q \notin A(i)$, then $|\mathcal{N}_q(i)| \geq 2$. Finally, if we use the last upper bound, we obtain that

$$\begin{aligned} \mathbf{C}_M(i) &\leq p c_G(i) \left(1 + 2(p - 1) + 2 \frac{(p - 1) \sum_{q=1}^p |\mathcal{N}_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \right) \\ &\leq p c_G(i) \left(1 + (p - 1) \left(4 + \frac{\theta(i)}{p - \theta(i)} \right) \right). \end{aligned}$$

On the other hand, if M is homogeneous, then the same techniques as before show that $c_G(i) \leq (p - \theta(i)) \mathbf{C}_M(i)$. Since M is homogeneous, then $\mathcal{N}_q(i) = \mathcal{N}(i)$ and $\overline{E}_q(i) \subseteq \overline{E}(i)$ for every $1 \leq q \leq p$. Therefore we conclude that

$$\begin{aligned} \mathbf{C}_M(i) &= \frac{2 \sum_{q=1}^p |\overline{E}_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q(i)| (|\mathcal{N}_q(i)| - 1)} \\ &\leq \frac{2p |\overline{E}(i)|}{\sum_{q=1}^p |\mathcal{N}(i)| (|\mathcal{N}(i)| - 1)} = \frac{2 |\overline{E}(i)|}{|\mathcal{N}(i)| (|\mathcal{N}(i)| - 1)} = c_G(i). \end{aligned}$$

□

Remark 2.5 The definition of $\theta(i)$ shows that the estimation in the previous theorem gets coarser as the number of slices for which i has strictly less than two neighbors (possibly zero) increases; in a “normal” scenario this situation will not occur and the estimation will involve a polynomial of degree two in p . Anyway, since $0 \leq \theta(i) \leq p$ and the function $f(x) = x/(p - x)$ is increasing in $[0, p)$, then the last result ensures that

$$\mathbf{C}_M(i) \leq g(p) c_G(i),$$

where $g(p)$ is of order p^α with $\alpha \in [2, 3]$, but still it is not obvious whether the exponent 2 can be lowered.

As it was said above there are other possible definitions for the clustering of a node in a multilevel network, based on the clusterings of that node in the slices, and then it is natural to have then related by means of estimations. However a remark is in order. Notice that it is possible to construct a network whose slices have clustering coefficients zero for every node while the clustering coefficient of the projection network is 1 for every node. Indeed, if we consider the example 2.3 as our multilevel network M , then it is straightforward to check that for a given node i ,

its clustering coefficient in each of the slices, $c_q(i)$, is zero, and hence every convex combination of them, while the clustering coefficient $\mathbf{C}_M(i)$ is 1. This means that we cannot expect to estimate $\mathbf{C}_M(i)$ from convex combinations of $\{c_q(i)\}_{q=1}^p$. Still it makes sense to give a “*slicewise*” definition for the clustering coefficient of the node i in a multilevel network M and find estimations by means of the clustering already defined. Again some notation is required. Call

$$\mathcal{N}_q^*(i) = \{j \in X \mid j \text{ is a neighbor of } i \text{ in } S_q\}$$

and $G_q(i) = (\mathcal{N}_q^*(i), E_q(i))$, where

$$E_q(i) = \{\{k, j\} \in E_q \mid k, j \in \mathcal{N}_q^*(i)\}$$

Note that $G_q(i)$ is the subgraph of the slice S_q generated by $\mathcal{N}_q^*(i)$. Thus the following definition is proposed.

Definition 2.6 Let $M = (X, E, \mathcal{S})$ be a multilevel network and i be a given node of M . The *slice clustering* coefficient of i is defined as

$$\mathbf{C}_G^{sl}(i) = \frac{2 \sum_{q=1}^p |E_q(i)|}{\sum_{q=1}^p |\mathcal{N}_q^*(i)| (|\mathcal{N}_q^*(i)| - 1)}.$$

Notice that $G_q(i)$ is a subgraph of $\overline{G}_q(i)$ as was illustrated in the example above. Accordingly $\mathcal{N}_q^*(i) \subseteq \mathcal{N}_q(i)$ and hence the largest possible number of links between neighbours of i in the q -slice, cannot exceed the corresponding largest possible number of links between neighbours of i in $\mathcal{N}_q(i)$. As above we have the relation

$$\frac{|\mathcal{N}(i)| (|\mathcal{N}(i)| - 1)}{2} \leq \frac{1}{2} \sum_{q=1}^p |\mathcal{N}_q^*(i)| (|\mathcal{N}_q^*(i)| - 1) + \sum_{k < j} |\mathcal{N}_k^*(i)| |\mathcal{N}_j^*(i)|,$$

where the last sum has $\binom{p}{2}$ terms. Also we can assume, after rearrangement, that $|\mathcal{N}_q^*(i)| \leq |\mathcal{N}_{q+1}^*(i)|$ for all $1 \leq q \leq p$. Thus by mimicking the proof of Theorem 2.4, we can get a relationship between the clustering coefficient in the projected network $c_G(i)$ and the slice clustering coefficient $\mathbf{C}_G^{sl}(i)$.

Theorem 2.7 Let $M = (X, E, \mathcal{S})$ be a multilevel network and i be a given node of M . Call $A(i) = \{q \in \{1, \dots, p\} \mid |\mathcal{N}_q^*(i)| < 2\}$ and $\theta(i) = |A(i)|$. Then

$$\mathbf{C}_G^{sl}(i) \leq p c_G(i) \left(1 + (p - 1) \left(4 + \frac{\theta(i)}{p - \theta(i)} \right) \right).$$

Notice that there are other possible variations for the definition of the slice clustering coefficient of a node i . For example, the average over the clustering coefficients $c_q(i)$ of the slides:

$$\mathbf{C}_G^{sl}(i) = \frac{1}{p} \sum_{q=1}^p c_q(i).$$

The relation between this definition and definition 2.6

$$p \min_{1 \leq q \leq p} \frac{|\mathcal{N}_q^*(i)|(|\mathcal{N}_q^*(i)| - 1)}{2} \mathbf{C}_G^{sl}(i) \leq \mathbf{C}_G^{sl}(i) \leq p \max_{1 \leq q \leq p} \frac{|\mathcal{N}_q^*(i)|(|\mathcal{N}_q^*(i)| - 1)}{2} \mathbf{C}_G^{sl}(i)$$

is easily derived from the following lemma.

Lemma 2.8 *If $a_1, a_2, \dots, a_n \in [0, \infty)$ and $b_1, b_2, \dots, b_n \in (0, \infty)$, then*

$$\frac{n \min_{1 \leq i \leq n} b_i}{\sum_{i=1}^n b_i} \left(\frac{1}{n} \sum_{i=1}^n \frac{a_i}{b_i} \right) \leq \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \leq \frac{n \max_{1 \leq i \leq n} b_i}{\sum_{i=1}^n b_i} \left(\frac{1}{n} \sum_{i=1}^n \frac{a_i}{b_i} \right).$$

Proof:

It is straightforward since

$$\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} = \sum_{i=1}^n \left(\frac{b_i}{\sum_{j=1}^n b_j} \right) \frac{a_i}{b_i}.$$

□

The following examples illustrate the behavior of the different concepts of clustering previously introduced. As it can be seen, if each one is separately evaluated one gets extremely different values.

Example 2.9 Consider $M = (X, E, \mathcal{S})$ with $X = \{1, 2, \dots, n\}$ and the slice set $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2\}$, where

$$\begin{aligned} \mathcal{S}_1 &= (\{1, 2, 3\}, \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}) \\ \mathcal{S}_2 &= (\{1, 2, \dots, n\}, \{\{1, 2\}, \{1, 3\}, \dots, \{1, n\}, \{2, 3\}\}). \end{aligned}$$

In this case

$$\mathbf{C}_G^{sl}(1) = \frac{1}{2} \left(\frac{1}{1} + \frac{1}{(n-1)(n-2)/2} \right) \sim \frac{1}{2}$$

while

$$c_G^{sl}(1) = \frac{1+1}{1+(n-1)(n-2)/2} = O\left(\frac{1}{n^2}\right).$$

We see that the two local clustering coefficients dramatically differ for $n \gg 1$.

On the other hand, if we take $M = (X, E, \mathcal{S})$ with $X = \{1, 2, \dots, n^2\}$ and the slice set is given by $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$, where $S_j = (X, E_j)$ and

$$\begin{aligned} E_1 &= \{\{i, j\} : 1 \leq i < j \leq n^2\}, \\ E_j &= \{\{1, 2\}, \dots, \{1, n\}, \{j, j+1\}\} \text{ if } 2 \leq j \leq n-1, \\ E_n &= \{\{1, 2\}, \dots, \{1, n\}, \{2, n\}\}, \end{aligned}$$

then

$$\begin{aligned} \mathbf{C}_G^{sl}(1) &= \frac{1}{n} \left(\frac{(n^2-1)(n^2-2)/2}{(n^2-1)(n^2-2)/2} + \frac{1}{(n-1)(n-2)/2} + \dots + \frac{1}{(n-1)(n-2)/2} \right) \\ &= O\left(\frac{1}{n^2}\right), \end{aligned}$$

while

$$\mathbf{C}_G^{sl}(1) = \frac{(n^2-1)(n^2-2)/2 + 1 + \dots + 1}{(n^2-1)(n^2-2)/2 + (n-1)(n-2)/2 + \dots + (n-1)(n-2)/2} \sim 1.$$

Also in this case the two local clustering coefficients are very different for $n \gg 1$.

Example 2.10 Consider $M = (X, E, \mathcal{S})$, where the slice $S_j = (X_j, E_j)$ is defined by $X_j = \{1, j+1, j+2\}$ and $E_j = \{\{1, j+1\}, \{1, j+2\}, \{j+1, j+2\}\}$ for $1 \leq j \leq n-2$, $X_{n-1} = \{1, 2, n\}$, $E_{n-1} = \{\{1, 2\}, \{1, n\}, \{2, n\}\}$. Then $\mathbf{C}_G^{sl}(1) = 1$ while $c_G(1) = O(1/n)$.

Example 2.11 Now consider $M = (X, E, \mathcal{S})$ with $X_j = X = \{1, 2, \dots, n\}$ and $E_j = \{\{1, k\} : 2 \leq k \leq n\} \cup \{\{j+1, k\} : 2 \leq k \leq n, k \neq j+1\}$ for $1 \leq j \leq n-1$. Then $c_G(1) = 1$ while $\mathbf{C}_G^{sl}(1) = O(1/n)$.

Example 2.12 If we take $M = (X, E, \mathcal{S})$ with $X_j = X = \{1, 2, \dots, n\}$ and $E_j = \{\{1, j+1\}, \{1, j+2\}, \{j+1, j+2\}\}$ for $1 \leq j \leq n-2$, $E_{n-1} = \{\{1, 2\}, \{1, n\}, \{2, n\}\}$. Then $\mathbf{C}_G^{sl}(1) = 1$ while $\mathbf{C}_M(1) = O(1/n^2)$.

Example 2.13 Consider $M = (X, E, \mathcal{S})$ with $X_j = X = \{1, 2, \dots, 2n\}$, $A = \{\{1, k\} : 2 \leq k \leq n\} \cup \{\{k, l\} : n+1 \leq k < l \leq 2n\}$, $E_j = A \cup \{\{j+1, j+2\}\}$ for $1 \leq j \leq n-2$, $E_{n-1} = A \cup \{\{n, 2\}\}$, $E_n = \{\{1, k\} : 2 \leq k \leq 2n\}$. Then $\mathbf{C}_M(1) \sim 1/4$ while $\mathbf{C}_G^{sl}(1) = O(1/n^2)$.

Example 2.14 Consider $M = (X, E, \mathcal{S})$ with $X_{i,j} = X = \{1, 2, \dots, n\}$ and $E_{i,j} = \{\{i, j\}\}$ for $1 \leq i < j \leq n$. Then $\mathbf{C}_G^{sl}(1) = 0$, $\mathbf{C}_M(1) = O(1/n^2)$ and $c_G(1) = 1$.

Example 2.15 Consider $M = (X, E, \mathcal{S})$ with $X = \{1, 2, \dots, n\}$, $X_1 = \{1, 2, 3\}$, with $E_1 = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ and $X_j = \{1, j+2\}$ with $E_j = \{\{1, j+2\}\}$ for $j = 2, \dots, n-2$.

Then

$$c_G(1) = \frac{1}{(n-1)(n-2)} = O(1/n^2)$$

while

$$\mathbf{C}_M(1) = \frac{1 + 0 + \dots + 0}{1 + 0 + \dots + 0} = 1.$$

We conclude this section with a final observation. All what has been said refers to local clustering coefficients. The aggregation of the local quantities into a global one, or clustering of M , can naturally be considered. We **leave** the details to the reader.

2.2 Metric Structure: Distances and Efficiency

If we want to introduce metric tools in a multilevel network $M = (X, E, \mathcal{S})$, we first have to give the notion of *path* and *length*. A *path* ω in $M = (X, E, \mathcal{S})$ is a set of the form $\omega = \{(\ell_1, \dots, \ell_q), (S_1, \dots, S_q)\}$ such that

- (i) $(\ell_1, \dots, \ell_q)$ is a path in (X, E) ,
- (ii) $(S_1, \dots, S_q)$ is a sequence of slice graphs $S_1, \dots, S_q \in \mathcal{S}$,
- (iii) For every $1 \leq j \leq q$, we have that ℓ_j is an edge in the slice graph S_j .

By using this concept we can introduce a metric structure in a multilevel graph $M = (X, E, \mathcal{S})$ as follows.

Definition 2.16 Let $M = (X, E, \mathcal{S})$ be a multilevel network, $\beta \geq 0$ fixed and $\omega = \{(\ell_1, \dots, \ell_q), (S_1, \dots, S_q)\}$ be a path in M . The *length* of ω is the nonnegative value

$$\ell(\omega) = q + \beta \sum_{j=2}^q \Delta(j),$$

where

$$\Delta(j) = \begin{cases} 1 & \text{if } S_j \neq S_{j-1}, \\ 0 & \text{otherwise.} \end{cases}$$

The *distance* in M between two nodes $i, j \in X$ is the minimal length among all possible paths in M from i to j .

If we take $\beta = 0$, the previous definition gives the natural metric in the projection graph, while if $\beta > 0$, we introduce new metrics that take into account, not only the global structure of the projection network, but also the interplay between the slice networks, that help to model the multi-scale nature of real-life social networks.

Remark 2.17 We could introduce a more general definition of the length of a path ω in a multilevel network $M = (X, E, \mathcal{S})$ of slices $S_1, \dots, S_p$ by replacing the *jumping weight* β by a $p \times p$ non-negative matrix $\Theta = (\beta(S_i, S_j))$ in order to take into account that some slices could be closer to an other and therefore we should take different *jumping weights* depending on the starting and target slices where the slice jump occurred. Following this idea, the length of a path $\omega = \{(\ell_1, \dots, \ell_q), (S_{\sigma(1)}, \dots, S_{\sigma(q)})\}$ (where $\sigma : \{1, \dots, q\} \rightarrow \{1, \dots, p\}$) could be defined as

$$\tilde{\ell}(\omega) = q + \sum_{j=2}^q \tilde{\Delta}(j),$$

where

$$\tilde{\Delta}(j) = \begin{cases} \beta(S_{\sigma(j-1)}, S_{\sigma(j)}) & \text{if } S_{\sigma(j)} \neq S_{\sigma(j-1)}, \\ 0 & \text{otherwise.} \end{cases}$$

We could also go further and consider that the *jumping weights* depend not only on the two slices involved in the slice jump, but also on the actual node in which this jump take place, but in any case, it is straightforward to check that the results, methods and algorithms introduced in this section can be easily mimicked to these more general metric structures.

One way to calculate the distance matrix, that is a matrix $\Lambda = (\lambda_{ij})$ that contains in each position λ_{ij} the distance in M between vertices i and j in X , is to consider an auxiliary graph in the following way. Every vertex of the multilevel network M is represented by a vertex in the auxiliary graph and, if a vertex in M belongs to two or more slice graphs in M , then we duplicated it as many times as the number of slice graphs it belongs to. Every edge of E is an edge in the auxiliary graph and there is one more (weighted) edge for each vertex duplication between the duplicated vertex and the original one. The distance between a duplicated vertex j and another vertex k is $d_{jk} = \min\{d_{j'k}, d_{j''k}\}$, where j' and j'' are the duplications of vertex j in the auxiliary graph.

Once we have defined a distance in a multilevel network $M = (X, E, \mathcal{S})$, it is easy to generalize the concept of efficiency given in [9] to this structure:

Definition 2.18 Let $M = (X, E, \mathcal{S})$ a multilevel network. The efficiency of M is

$$\mathbf{E}(M) = \frac{1}{n(n-1)} \sum_{i,j \in X, i \neq j} \frac{1}{d_M(i, j)},$$

where n is the number of vertices in X and $d_M(i, j)$ is the distance in M between vertices i and j .

It is quite natural to try to establish comparisons between the efficiency $\mathbf{E}(M)$ of a multilevel graph $M = (X, E, \mathcal{S})$ as it was just defined before and the efficiency $E(G)$ of the underlying graph as a classic complex network, as it was introduced in [13] (see also [6]). The actual analytical result that connects these two parameters are given in the following theorem.

Theorem 2.19 Let $M = (X, E, \mathcal{S})$ a multilevel network, and $G = (X, E)$. Then

$$\frac{1}{\beta + 1}E(G) \leq \mathbf{E}(M) \leq E(G), \quad (2.6)$$

where $E(G)$ is the efficiency of the projection network G and $\mathbf{E}(M)$ is the efficiency of the multilevel network M .

Proof:

Let $i, j \in X$ and ω be a path in M from i to j . We denote by $d_M(i, j)$ the distance between i, j in the multilevel network M and by $d_G(i, j)$ the distance between i, j in the network G . In a similar way, we denote by $\ell_M(\omega)$ its length in the multilevel network M and by $\ell_G(\omega)$ its length in the network G .

It is clear that $d_M(i, j) \geq d_G(i, j)$. On the other hand, since $\Delta(\cdot) \leq 1$ then

$$\begin{aligned} d_M(i, j) &\leq \ell_M(\omega) = \ell_G(\omega) + \beta \sum_{j=2}^{\ell_G(\omega)} \Delta(j) \\ &\leq \ell_G(\omega) + \beta(\ell_G(\omega) - 1) = \ell_G(\omega)(\beta + 1) - \beta \leq \ell_G(\omega)(\beta + 1), \end{aligned}$$

Hence, since it is clear that $d_G(i, j) = \min\{\ell_G(\omega) | \omega \text{ is a path in } M \text{ from } i \text{ to } j\}$, we get that

$$\frac{1}{\beta + 1}d_M(i, j) \leq d_G(i, j)$$

and therefore, by combining the last expressions with the reverse bound obtained before we obtain that

$$\frac{1}{\beta + 1} \frac{1}{d_G(i, j)} \leq \frac{1}{d_M(i, j)} \leq \frac{1}{d_G(i, j)},$$

which gives the result. □

Remark 2.20 The last bounds are sharp simply by making $\beta \rightarrow 0^+$.

We are equally interested in obtaining estimations of the efficiency of a multilevel graph in terms of the efficiencies of its slices. Precisely we obtain the following result.

Theorem 2.21 If $M = (X, E, \mathcal{S})$ is a multilevel network such that $S = \{S_1, \dots, S_p\}$ and for every $1 \leq q \leq p$ we denote $S_q = (X_q, E_q)$, then

$$\frac{|X_q|(|X_q| - 1)}{|X|(|X| - 1)}E(S_q) \leq \mathbf{E}(M), \quad (2.7)$$

where $|X_q|$ is the number of nodes of S_q and $|X|$ is the number of nodes of M .

Proof:

If we take $1 \leq q \leq p$, since $X_q \subseteq X$ and $E_q \subseteq E$, then for every $i, j \in X_q$ and every path ω in S_q from i to j in S_q , we get that $\ell_{S_q}(\omega) = \ell_M(\omega) \geq d_M(i, j)$. Then $d_{S_q}(i, j) \geq d_M(i, j)$. Therefore

$$\sum_{i,j \in X_q, i \neq j} \frac{1}{d_{S_q}(i, j)} \leq \sum_{i,j \in X_q, i \neq j} \frac{1}{d_M(i, j)} \leq \sum_{i,j \in X, i \neq j} \frac{1}{d_M(i, j)},$$

which makes that

$$\begin{aligned} E(S_q) &= \frac{1}{|X_q|(|X_q| - 1)} \sum_{i,j \in X_q, i \neq j} \frac{1}{d_{S_q}(i, j)} \\ &\leq \frac{1}{|X_q|(|X_q| - 1)} \sum_{i,j \in X, i \neq j} \frac{1}{d_M(i, j)} = \frac{|X|(|X| - 1)}{|X_q|(|X_q| - 1)} \mathbf{E}(M). \end{aligned}$$

□

Remark 2.22 There is a reverse inequality that can be established when $M = (X, E, \mathcal{S})$ is an homogeneous multilevel network (i.e., a multilevel network such that $X_q = X$ for every slice $S_q = (X_q, E_q) \in \mathcal{S}$). Indeed, in this case the following relation is easily obtained

$$E(S_q) \leq \mathbf{E}(M) \leq E(G) \leq (|X| - 1)E(S_q). \quad (2.8)$$

3 Some models for multilevel random networks

In this final section we present some models to produce multilevel random networks, inspired by the Barabási-Albert preferential attachment model of complex networks [3] and several bipartite networks models such as the collaboration network model proposed by Ramasco *et al.* [17] or the sublineal preferential attachment bipartite model introduced by M. Peltomäki and M. Alava [16]. We propose growing random models since many of the real-life examples that can be modeled by multilevel networks, such as social or transportation networks, are dynamic models and their structure **grows** in time by addition of new nodes, links and layers. The presented randomized models are determined by the following rules:

- (i) *Model parameters.* Our models have three main parameters: N , m and p_{new} . We set $N \in \mathbb{N}$ as the number of nodes in the multilevel network while $2 \leq m \leq N$ account of the number of nodes in each layer (i.e. if we take $m = 2$, we recover the Barabási-Albert model [3]).

In our model m will be fixed, but it can also be replaced by any other non-negative integer random variable if we want to produce more general models. Finally, we set $p_{new} \in (0, 1]$ as the probability of joining a new node to the multilevel network.

- (ii) *Initial conditions.* We start with a seed multilevel network made of one single layer S_0 of m nodes that are linked all to all, (*i.e.* $S_0 = K_m$). We can replace the *all-to-all* structure by any other structure (such as a scale free or a Erdős-Rényi network), but the results obtained are statistically equivalent.
- (iii) *Layer composition.* At each time step t , a new layer S_t of m nodes is added to the multilevel network. In order to determine S_t we have to give its nodes and links. We start by choosing randomly an existing node of the multilevel network proportionally to its degree (*preferential election*). Therefore if at step $t - 1$, the set of nodes of the multilevel network is $\{v_1, \dots, v_n\}$, and k_i denotes the degree of node v_i at time $t - 1$ in the projection network, then we choose the node v_i randomly and independently with probability

$$p_i = \frac{k_i}{\sum_{j=1}^n k_j}.$$

The chosen node will be the first node of the new layer S_t and we call it *coordinator* of the layer. As an alternative, we could choose the coordinator proportionally to the number of layers that it belongs to but, in this case, the model will be similar to the collaboration network model proposed by Ramasco et al. [17]. Each of the remaining $m - 1$ nodes of S_t will be a new node with probability p_{new} and an existing node with probability $(1 - p_{new})$. If we have to add an already existing node, we will uniformly and independently choose it at random. Note that we can replace the uniform random selection by other random procedure (such as preferential selection), but the random tests done suggest that the multilevel network obtained have statistically the same structural properties when N is large enough ($N > 10^3$). Anyway, we should have chosen m nodes $\tilde{v}_1, \dots, \tilde{v}_m$ ($\tilde{v}_1$ is the coordinator node) that will belong to the new layer S_t .

- (iv) *Layer inner-structure.* Once we have fixed the nodes $\tilde{v}_1, \dots, \tilde{v}_m$ of the new layer S_t , we have to give its links. First, we link all the nodes to the coordinator node in order to warrant that the new layer is connected. In addition to this, we set new links between each pair of nodes, say v_i and v_j (with $1 < i \neq j \leq m$) by using a random linking probability p_{ij} that we will present later. At the end of this step we have defined completely the new layer S_t .
- (v) Finally, we repeat steps (iii) and (iv) until the number of nodes of the multilevel network is at least N .

These rules define a family of growing models for multilevel random networks that cover a wide range of different networks with different meso-scale structure, simply by changing the *linking strategy*, p_{ij} , at each step. The simplest choice for this strategy is $p_{ij} = 1$, *i.e.* an all-to all strategy, thus linking all the nodes in the new slice. In this way, we recover a model similar to the Ramasco *et al.* model [17],

that can be used to produce hyper-networks with a meso-scale structure similar to that found in scientific collaboration systems.

In this paper we will focus on the following three different linking strategies:

1. *Erdős-Rényi type strategy (Model I)*. We fix a value $p_{ij} = p_{link} \in [0, 1]$ so that we add each link $\{\tilde{v}_i, \tilde{v}_j\}$ randomly and independently with probability p_{link} , for every $2 \leq i \neq j \leq m$ (in what follows we will use a value $p_{link} = 0.31$). Note that this strategy does not take into account the meso-scale structure of the multilevel network, since the existence of the link $\{\tilde{v}_i, \tilde{v}_j\}$ is independent of the nodes $\tilde{v}_i, \tilde{v}_j$ in other slices of the multilevel network. Clearly, this is a toy model inspired by the classic Erdős-Rényi model.
2. *Assortative linking strategy (Model II)*. For every $2 \leq i \neq j \leq m$, we add randomly the link $\{\tilde{v}_i, \tilde{v}_j\}$ proportionally to the number of common slices that hold simultaneously $\tilde{v}_i$ and $\tilde{v}_j$. Hence if we denote by Q_{ij} the number of slices that hold simultaneously $\tilde{v}_i$ and $\tilde{v}_j$ at time step t (including S_t) and by q_i the number of slices that hold $\tilde{v}_i$ at time step t (also including S_t), thus the probability of linking node $\tilde{v}_i$ with node $\tilde{v}_j$ is given by

$$p_{ij} = \frac{2Q_{ij}}{q_i + q_j},$$

for every $2 \leq i \neq j \leq m$. The heuristic behind this strategy comes from social networks, since in this type of networks the relationships in a new social group are correlated with the previous relationships between the actors in other social groups [20]. Hence, on the one hand, if two actors that belong to the new social group coincide in many (previous) social groups, then the probability of linking in this new social group is big and, on the other hand, when two new actors join their first social group, the probability of establishing a relationship between them it is also high.

3. *Disassortative linking strategy (Model III)*. If we take $2 \leq i \neq j \leq m$, the probability of linking $\tilde{v}_i$ with $\tilde{v}_j$ in the slice S_t is inversely proportional to the number of slices that hold simultaneously $\tilde{v}_i$ and $\tilde{v}_j$, i.e. if we denote by p_{ij} such probability, then

$$p_{ij} = 1 - \frac{2Q_{ij}}{q_i + q_j},$$

by using the same notation as in the assortative strategy (Model II). Now, the heuristic inspiring this strategy lies in some transportation networks, such as the airline networks, where the links in a new line try to connect nodes that are not connected in previous lines. By using the last expression, the newcomer nodes of the slice will be linked to old nodes that belong to many slices with high probability, while the newcomers will not be linked between them also with high probability. This strategy prevents the multiple reiteration of the

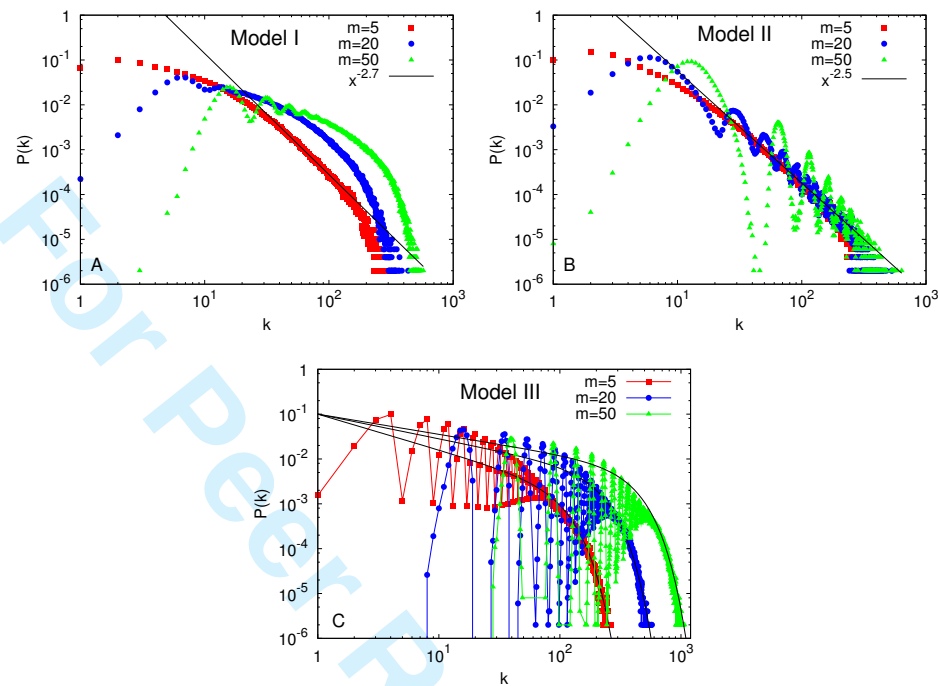


Figure 2: Random testing of the degree distribution of the projection graph in the randomized multilevel models. The model I is on the left, the model II is centered and the Model III is on the right, each of them for $m = 5, 20, 50$ nodes per slice. In all cases, the probability of joining a new node to the network is fixed to $p_{new} = 0.202$ as in the empirical data obtained in [16]. The distribution of each type of network and size of slice is computed for 100 random networks of 5000 nodes each one.

same link in different slices and stresses the dissasortativity between newcomer nodes and nodes that belong to many slices.

Once we have introduced the three linking strategies, we now present some numerical results to shows that, despite the fact that the above tree rules produce hyper-graphs with similar projection networks, the structures found in the mesoscale can be quite different.

Let us start by showing in figure 2 the degree distribution, $P(k)$, of the projection network. It is clear that the degree distribution of the projection network obtained in three models display heavy-tailed profiles, and this fact is more relevant if the number of nodes of each slice is low, since in this case the Big Number Law's make that the degree distribution behaves as a power-law, $P(k) \sim k^{-\gamma}$ with $2 \leq \gamma \leq 3$. Note that a discretization effect appears in all the distribution presented in figure 2 that makes that the degree distribution are oscillatory. This is due to the fact that each slice has the same number of nodes, and all of them are linked to the coordinator

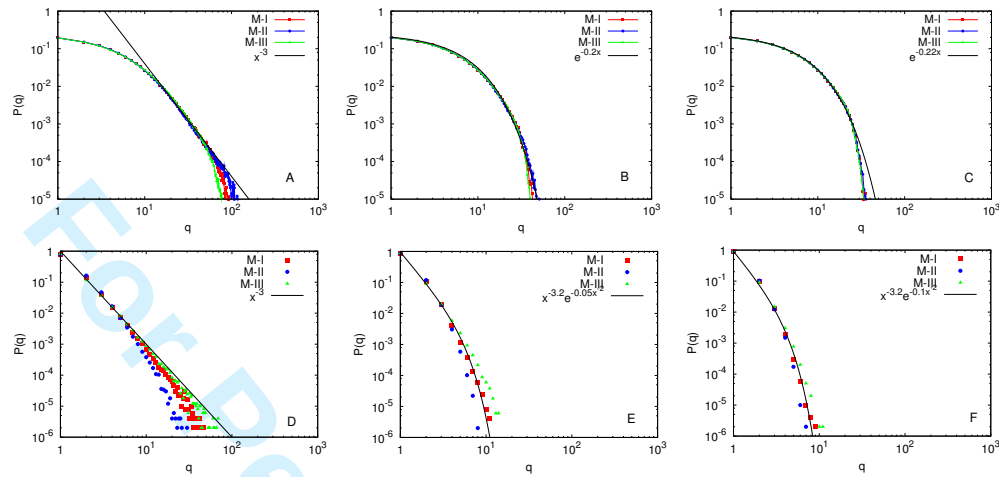


Figure 3: Random testing of the hyper-degree distribution, $P(q)$, of the multilevel models (in log scale) for different number of nodes m in each slice and for two different values of p_{new} , combined with some numerical approximation in each case. The cases $m = 5$ are on the left, $m = 20$ are centered and those for $m = 50$ are on the right. Besides, the panels on the top show the case $p_{new} = 0.202$ while those on the bottom are for $p_{new} = 0.9$. Each distribution $P(q)$ is averaged over 100 random networks of 5000 nodes each one.

node, which makes that the degree of the nodes are concentrated in some values modulus m . In order to avoid this effect, we could modify step (i) replacing m by a non-negative integer random variable that prevents that the degree of each node increases modulo m . This discretization effect is present in the three models and increases when the size of the slices is large, but it is more persistent in Model III, since in this case the dissasortative character of the model stresses this phenomenon.

In addition to the degree distribution of the projection network, in figure 3 we show the distribution $P(q)$ of the number of slices that hold each node (we call it *hyper-degree distribution* for consistency with the hyper-graph theory notation). Obviously, the hyper-degree of a node, q_i , depends on the number of nodes m of each slide and the probability p_{new} of adding a new node for constructing each slide (so that the number of new nodes in each slide is on average mp_{new}). Therefore, the distribution $P(q)$ is not correlated with the linking model used to construct the network. In figure 3 we show that the behavior of $P(q)$ depends quite strongly on the particular value of m and p_{new} . In particular, in figures 3.A, 3.B and 3.C we show respectively the cases $m = 5$, $m = 20$ and $m = 50$ when p_{new} is small ($p_{new} = 0.202$ as in figure 2). For $m = 5$, $P(q)$ can be approximated by a stretched exponential (figure 3.A), while $P(q)$ presents an exponential decay in the cases $m = 20$ and $m = 50$ (figures 3.B and 3.C). When p_{new} is large ($p_{new} = 0.9$ in figures 3.D, 3.E and 3.F) the hyper-degree distributions become more heterogeneous as they are better

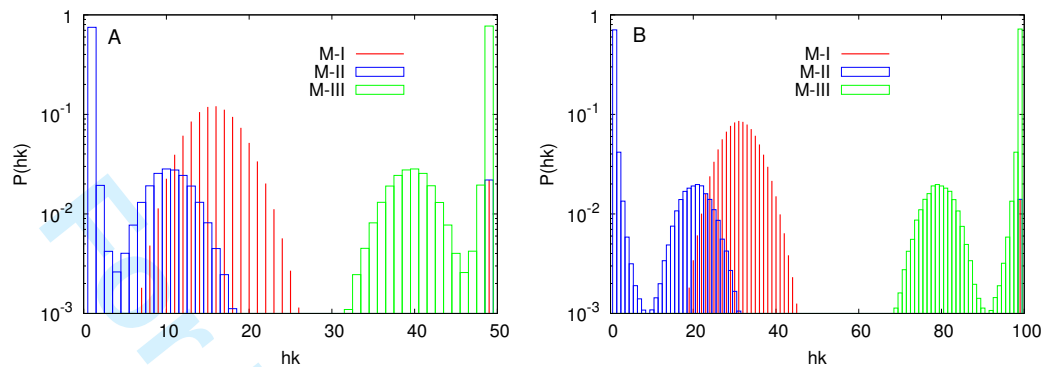


Figure 4: Random testing of the average degree distribution of each slice for 100 random networks of 5000 nodes each one in the randomized multilevel models I, II and III. The case $m = 50$ is in panel A (on the left), while the case $m = 100$ is in panel B (on the right), both for models I, II and III.

approximated by truncated power-laws. Remarkably, for $m = 5$ (figure 3.D) we obtain the most heterogeneous distribution since $P(q) \sim q^{-3}$ for nearly all the range of q values. The heterogeneity of $P(q)$ decreases with m as the plots for $m = 20$ and $m = 50$ (figures 3.E and 3.F) show. In these latter cases the power law behavior spans the small hyper-degree classes while, for large q values, a cut-off of the form $\exp(-\alpha q^2)$ applies.

Figures 2 and 3 have shown that either structure of the projection network and hyper-degree distribution are quite similar for the three random models presented. However, they are very different if we analyze their mesoscale structure, as figure 4 shows. This figure shows, for $m = 50$ and $m = 100$, the probability of finding a node connected to hk nodes within a randomly chosen slice of the network. It is straightforward to check that since the Erdős-Rényi type strategy is used in Model I, then the resulting distribution $P(hk)$ is a binomial distribution with parameters m and p_{link} (asymptotically a Poisson distribution). However, note that in figure 2 we found that the degree distribution of the projection network is scale-free. Correspondingly, Models II and III display totally different mesoscale properties as shown by their corresponding distributions $P(hk)$. On one hand in Model II we observe that $P(hk)$ tend to accumulate around the maximum possible value of hk , m , due to the assortative character of the linking strategy used. Conversely, in Model III the association between the nodes belonging to the same slice is much lower due to the disassortativity of the model. Again, while both Models II and III gives two different bimodal averaged distributions $P(hk)$, the corresponding distribution for their projection networks are also of scale-free type. Therefore, these results illustrate that many different multilevel networks (with different mesoscale structures) can produce similar projection networks and therefore that we should take into account

the mesoscale structure in order to give sharper models of many real-life problems.

Acknowledgements

The authors of this work have been partially supported by the Spanish Government Project MTM2009-13848.

References

- [1] R. ALBERT AND A.L. BARABÁSI, *Statistical mechanics of complex networks*, Rev. Mod. Phys. **74** (2002), 47–97.
- [2] A.L. BARABÁSI, *Linked: The new science of networks*, Perseus Publishing, Cambridge, Massachusetts, 2002.
- [3] A.L. BARABÁSI AND R. ALBERT, *Emergence of scaling in random networks*, Science **286**, 509–512.
- [4] C. BERGE, *Hyper-graphs. Combinatorics of Finite Sets*, North-Holland, 1989.
- [5] S. BOCCALETTI, V. LATORA, Y. MORENO, M. CHAVEZ AND D.U. HWANG, *Complex Networks: Structure and dynamics*, Physics Reports **424** (2006), 175–308.
- [6] R. CRIADO, A. GARCÍA DEL AMO, B. HERNÁNDEZ-BERMEJO AND M. ROMANCE, *New Results on computable efficiency and its stability for complex networks*, J. Comput. Appl. Math. **292** (1) (2006), 059–074.
- [7] R. CRIADO, J. FLORES, M.I. GONZÁLEZ-VASCO AND J. PELLO, *Choosing a leader on a complex network*, J. Comput. Appl. Math. **204** (1) (2007), 010–017.
- [8] R. CRIADO, J. PELLO, M. ROMANCE AND M. VELA-PÉREZ, *Improvements in performance and security for complex networks*, Int. J. Comput. Math. **86** (2) (2009), 209–218.
- [9] R. CRIADO, M. ROMANCE AND M. VELA-PÉREZ, *Hyperstructures, a new approach to complex systems*, Int. J. Bif. Chaos **20** (2010), 877–883.
- [10] E. ESTRADA AND J.A. RODRÍGUEZ-VELÁZQUEZ, *Subgraph centrality and clustering in complex hyper-networks*, Physica A (2006).
- [11] M. GIRVAN AND M.E.J. NEWMAN, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. USA **99** (2002), 7821.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

[12] B. GUERRA, J. PONCELA J. GÓMEZ-GARDEÑES, V. LATORA AND Y. MORENO, *Dynamical organization towards consensus in Axelrod model on complex networks*, Phys. Rev. E **81** (2010), 056105.

[13] V. LATORA, M. MARCHIORI, *Efficient Behavior of Small-World Networks*, Phys. Rev. Lett. **87** (2001),198701.

[14] M.E.J. NEWMAN, *The structure and function of complex networks*, SIAM Review **45** (2003), 167–256.

[15] M.E.J. NEWMAN, *Detecting community structure in networks*, Eur.Phys.J.B **38** (2004), 321–330.

[16] M. PELTOMÄKI AND M. ALAVA *Correlations in bipartite collaboration networks*, J. Stat. Mech. (2006), P01010.

[17] J.J. RAMASCO, S.N. DOROGVTZEV AND R. PASTOR-SATORRAS, *Self-organization of collaboration networks*, Phys. Rev. E **70** (2004), 036106.

[18] J. SCOTT, *Social Networks Analysis: A Handbook*, 2nd ed., Sage, London, 2000.

[19] S.H. STROGATZ, *Exploring complex networks*, Nature **410** (2001), 268–276.

[20] S. WASSERMAN AND K. FAUST, *Social Networks Analysis*, Cambridge University Press, Cambridge, 1994.

[21] D.J. WATTS, D.J., STROGATZ, S.H., "Collective dynamics of small-world networks", *Nature* **393**, 440–442 (1998).