

RESEARCH

Open Access



# The structure of co-publications multilayer network

Ghislain Romaric Meleu<sup>1,2,3\*</sup> and Paulin Yonta Melatagia<sup>2,3</sup>

\*Correspondence:  
meleu.romaric@gmail.com

<sup>1</sup> Département de  
Mathématiques et  
Informatique, Faculté  
des Sciences, Université  
de Ngaoundéré, B. P.  
454 Ngaoundéré, Cameroun  
Full list of author information  
is available at the end of the  
article

## Abstract

Using the headers of scientific papers, we have built multilayer networks of entities involved in research namely: authors, laboratories, and institutions. We have analyzed some properties of such networks built from data extracted from the HAL archives and found that the network at each layer is a small-world network with power law distribution. In order to simulate such co-publication network, we propose a multilayer network generation model based on the formation of cliques at each layer and the affiliation of each new node to the higher layers. The clique is built from new and existing nodes selected using preferential attachment. We also show that, the degree distribution of generated layers follows a power law. From the simulations of our model, we show that the generated multilayer networks reproduce the studied properties of co-publication networks.

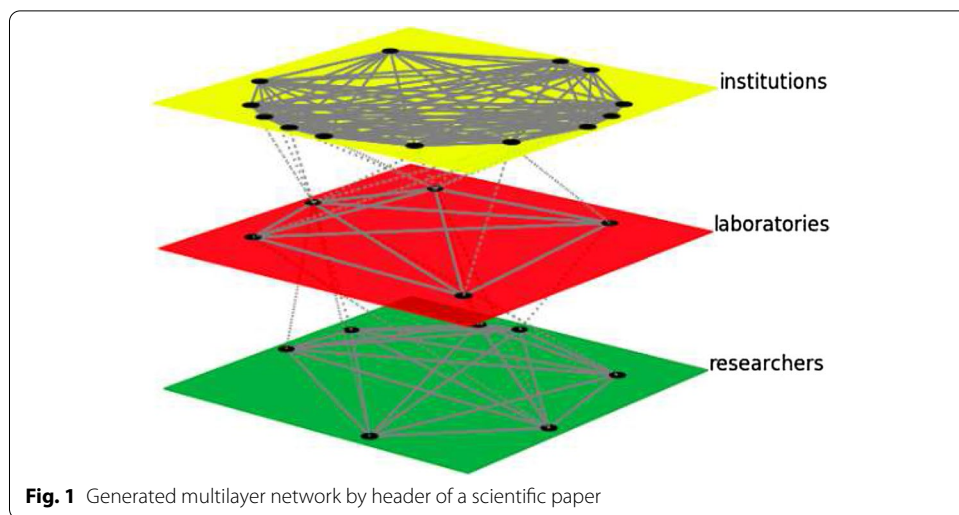
**Keywords:** Multilayer networks, Collaboration networks, Clique, Preferential attachment

## Introduction

The recent years of research in networks science have been characterized by many and more attempts to generalize the traditional network theory by developing and validating a novel framework for the study of multilayer networks, i.e., graphs where constituents are nodes, and several layers of connections have to be taken into account to accurately describe the nodes interactions. Multilayer networks explicitly incorporate multiple channels of connectivity and constitute the natural environment to describe systems interconnected through different categories of connections: each channel (relationship, activity, category, etc.) is represented by a layer and the same node or entity may have different kinds of interactions (different set of neighbors in each layer). For instance, in social networks, one can consider several types of different relationships: friendship, vicinity, kinship, membership of the same cultural society, partnership or coworker-ship, etc [1].

Such a change of paradigm that was termed in disparate ways (multiplex networks, networks of networks, interdependent networks, hypergraphs, and many others), already led to a series of very relevant and unexpected results [1].

Using the headers of scientific papers, we have represented and analyzed three networks: researchers, laboratories and institutions networks. Header of a paper is the



name of the researchers(authors) and their affiliations. The term institution is used here to refer to an university, a research center or a research institution. The networks are interdependent because of the affiliation relationship between the entities of the three networks. In fact, a researcher is affiliated to at least one laboratory and a laboratory affiliated to at least an institution. We therefore say that these networks are hierarchical and are induced from collaboration between researchers (see Fig. 1). In order to understand and simulate these systems we have to master both how the actors interact and the rules of affiliations to organizations.

We generalize this hierarchical co-publication network by a complex system where an actor is affiliated to an organization that can also be affiliated to a higher level organization and so on. The relationships between entities at the same level are deduced from interaction of those at the lower level. The interaction between the actors (at the lowest level) is the process that induces the relationship between the organization at the other levels. i. e, two researchers are not connected because they are members of the same laboratory but because they co-published a paper. This is the main difference between affiliation network used in this paper and those of Silvio Lattanzi et al. [2] who consider that in social networks there are two types of entities, actors and societies, that are linked by affiliation of the former to the latter.

After some definitions and the state of the art presented in "[Multilayer networks generation models](#)" section, the method applied for the work conducted in this paper was firstly to use the headers of 70224 scientific papers of the HAL archives to build co-publications networks. We measured various structural properties of these networks such as degree-distributions, average distance, clustering coefficient, ... "[Co-publications multilayer networks](#)" section is dedicated to this first part of our contribution. From this study, we derived a hierarchical network generation model that can reproduce the measured properties. The model is presented in "[The hierarchical network generation model](#)" section with some mathematical results regarding the number of nodes, the number of edges and exponent of the power law degree-distribution of the generated networks. Finally in "[Simulations](#)" section, we present some simulations results of the proposed generation model and discussions concerning the comparison between the properties of

the simulated networks and those of the real-world networks (built from the HAL dataset). "Conclusion" section concludes the paper.

### Multilayer networks generation models

The last 15 years have seen the birth of a movement in science: the complex networks theory. It involved the interdisciplinary effort of some of our best scientists in the aim of exploiting the current availability of big data in order to extract the ultimate and optimal representation of the underlying complex systems and mechanisms. The main goals were (i) the extraction of unifying principles that could encompass and describe (under some generic and universal rules) the structural accommodation that is being detected ubiquitously, and (ii) the modeling of the resulting emergent dynamics to explain what is actually seen and experienced from the observation of such systems.

Network theory provides various tools for investigating the structural or functional topology of many complex systems found in nature, technology and society. There are many applications of multilayer graphs in various areas such as biology, transportation and social network [3–7].

### Definitions and notations

A network (or graph) is a pair  $G = (X, E)$ , where  $X$  set of items, which we will call nodes and  $E$  is a set of connections between the nodes, called edges. A set of nodes joined by edges is only the simplest type of network; there are many ways in which networks may be more complex than this. For instance, there may be more than one different type of node in a network or more than one different type of edge. Nodes or edges may also have a variety of properties, numerical or categorical associated.

Graphs of directed edges [8] are themselves called directed graphs or sometimes digraphs. One can also have hyperedges—edges that join more than two nodes together. Graphs containing such edges are called hypergraphs [9]. Graphs may also be naturally partitioned in various ways. For example bipartite graphs: graphs that contain nodes of two distinct types, with edges running only between unlike types [10, 11]. So-called affiliation networks in which people are joined together by common membership of groups take this form, the two types of nodes representing the people and the groups [2].

There is no consensual definition of multilayer graphs. There are several definitional approaches in the literature [1, 12, 13]. In this work, we will use the definition of [1]. A multilayer network is a pair  $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ , where  $\mathcal{G} = \{G_\alpha; \alpha \in \{1, \dots, M\}\}$  is a family of (directed or undirected, weighted or unweighted) graphs  $G_\alpha = (X_\alpha, E_\alpha)$  (called layers of  $\mathcal{M}$ ) and

$$\mathcal{C} = \{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta; \alpha, \beta \in \{1, \dots, M\}, \alpha \neq \beta\}$$

is the set of interconnections between nodes of different layers  $G_\alpha$  and  $G_\beta$  with  $\alpha \neq \beta$ . The elements of  $\mathcal{C}$  are called crossed layers, and the elements of each  $E_\alpha$  are called intra-layer connections of  $\mathcal{M}$  in contrast with the elements of each  $E_{\alpha\beta}$  ( $\alpha \neq \beta$ ) that are called interlayer connections.

This mathematical model is suited to describe phenomena in social systems, as well as many other complex systems. By using this representation, we simultaneously take into

account: the links inside the different groups, the nature of the links and the relationships between elements that (possibly) belong to different layers and the specific nodes belonging to each layer involved.

It is important to notice that the concept of multilayer network extends that of other mathematical objects, such as: Multiplex network [14], Temporal networks [15], Interacting or interconnected networks [16], Multidimensional networks [17], Interdependent (or layered) networks [18, 19], Multilevel networks [20], Hypernetworks (or hypergraphs) [9].

The degree of a node [21]  $i \in X$  of a multiplex network  $\mathcal{M} = (\mathcal{G}, \mathcal{C})$  is the vector  $k_i = [k_i^1, \dots, k_i^M]$ , where  $k_i^\alpha$  is the degree of the node  $i$  in the layer  $\alpha$ . This vector-type node degree is the natural extension of the established definition of the node degree in a monolayer network.

We say that node  $i$ , with  $i = 1, 2, \dots, N$  is active at layer  $\alpha$  if  $k_i^\alpha > 0$ . We can then associate to each node  $i$  a node-activity vector  $b_i = \{b_i^{[1]}, b_i^{[2]}, \dots, b_i^{[M]}\}$ , where  $b_i^{[\alpha]} = 1$  if  $k_i^\alpha > 0$ , while  $b_i^{[\alpha]} = 0$  otherwise. We call node-activity  $B_i$  of node  $i$  the number of layers on which node  $i$  is active.

### Properties of real-world complex networks

It has been recently shown that most real-world complex networks have some essential properties in common. Three properties received much attention due to the fact that they have unexpected behaviors in real-world complex networks: the average distance between nodes, the clustering and the degree distribution.

Most of real-world complex networks have the small-world property, i.e., short average distance [22, 23]. The small-world concept originated from the famous experience made by Milgram [24]. Another property of many real-world networks is the presence of high average clustering coefficient.

The degree distribution which is, for each  $k$ , the probability  $p_k$  that a randomly chosen node has degree  $k$ , is completely different from what was expected. Indeed, for almost all real-world complex networks, the degree distribution follows a power law:  $p_k \approx k^{-\alpha}$ . The exponent  $\alpha$  of the power law is generally between 2 and 3. Such a distribution means that although most nodes have a small degree, the number of nodes with degree  $k$  decays only polynomially with  $k$ , and therefore there is a significant number of nodes with high degree. It has been shown in the literature that many co-authorship networks follow power law degree distribution [25–27].

### The state of the art of generation models

There are basically two ways to propose a model for network generations:

- The first may consider a set of observed properties as essential, and then sample randomly objects among the ones which have these properties. Proceeding this way, will yields a typical object with the concerned properties [28–30]. It is then possible to determine if the retained set of properties is sufficient (do the random objects produced by the model fit well the real one? ) and to study the expected behavior of the object of interest. The relevance of the set of properties is generally checked using other known properties or behaviors of the object.

- The second define's a construction network generation models process inspired from the way the object is really constructed [2, 31–33]. This construction process is generally iterated from an initial state, and eventually leads to an appropriate object. The analysis then concerns the properties induced by the construction process: do they fit real-world properties?

For more details, the reader can refer to [34] for a large overview on the model of simple network. Similarly to monolayer networks, most of the models for generation of multilayer networks can be divided also into two classes:

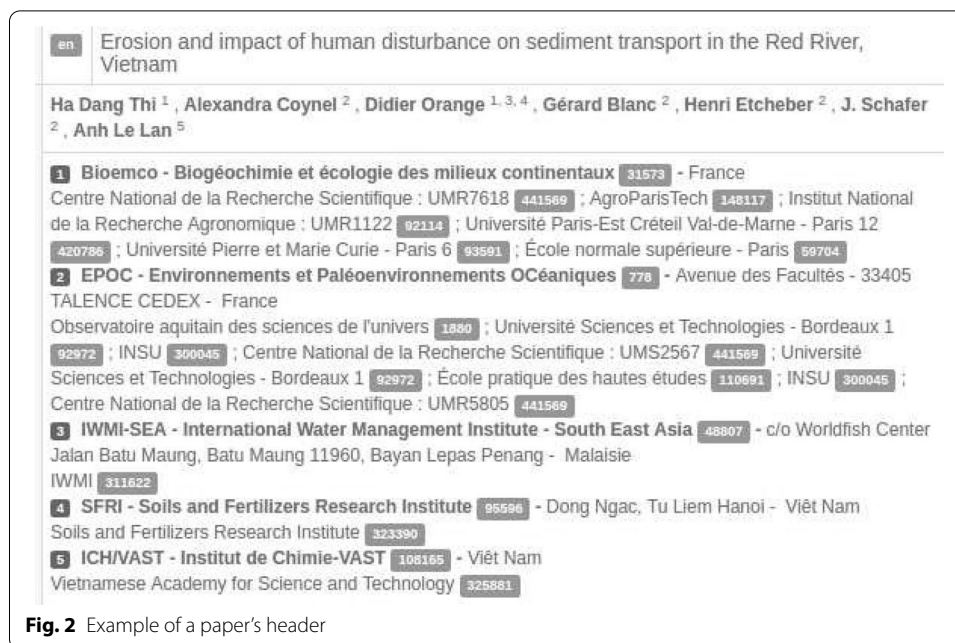
- Growing multilayer networks models, in which the number of the nodes grows, and there is a generalized preferential attachment rule [35–37]. These models explain multilayer network evolution starting from simple, and fundamental rules for their dynamics.
- Multilayer network ensembles, which are ensembles of networks with  $N$  nodes in each layer satisfying a certain set of structural constraints [38–40]. These ensembles are able to generate multilayer networks with fully controlled set of degree–degree correlations and of overlap.

In [33] Meleu and Melatagia proposed a networks generation model based on the formation of cliques to reproduce collaboration networks. At each step of the model, a clique of  $\lambda\eta$  existing nodes and  $(1 - \lambda)\eta$  new nodes is created and added in the network;  $P$  is the distribution of the number of nodes per collaboration,  $\eta$  is the mathematical expectation of  $P$  and  $\lambda$  is the proportion of old nodes per clique. The old nodes are selected according to preferential attachment. The main difference between the model of Zhang et al. [41] and those of Meleu et al. is that, Zhang et al. consider only one new node while Meleu et al. define a parameter  $\lambda$  that controls the proportion of new nodes. The model of Meleu et al. is thus the generalization of the model of Zhang et al. (see [33] for details).

In [35, 36] a growing multiplex model has been proposed: the network has a dynamics dictated by growth, and generalized preferential attachment. Starting at time  $t = 0$  from a duplex network with  $n_0$  nodes (with a replica in each of the two layers) connected by  $m_0 > m$  links in each layer, the model proceeds as follows:

- Growth: At each time  $t \geq 1$  a node with a replica node in each of the two layers is added to the multiplex. Each newly added replica node is connected to the other nodes of the same layer by  $m$  links.
- Generalized preferential attachment: The new link in layers  $\alpha = 1, 2$  is attached to node  $i$  with probability  $\Pi_i^\alpha$  proportional to a linear combination of the degree  $k_i^{[1]}$  of node  $i$  in layer 1 and  $k_i^{[2]}$  of node  $i$  in layer 2.

Growing multiplex network models have been proposed in [20, 42], where the multiplex network grows by the addition of an entire new layer at each step. In [20], two nodes  $i$  and  $j$  in the new layer are linked with a probability  $p_{ij}$  that depends on the quantity there called node *multiplexity*  $Q_{ij}$ . In particular,  $p_{ij}$  can be either positively correlated with  $Q_{ij}$ , or negatively correlated to it. In the first case two nodes that are active at the same time in many layers are more likely to be connected in the new layer, in the second case two



nodes that are active at the same time in many layers have small probability to be connected in the new layer. In [42], instead, every node  $i$  of the new layer will be active with a probability  $P_i$  proportional to the activity of the node  $B_i$ :  $P_i \approx A + B_i(t)$ , where  $A$  is a parameter of the model. This model enforces a sort of “preferential attachment” of the new layers to nodes of high activity  $B_i$ , and a power law distribution  $P(B_i)$  of the activities of the nodes.

The simplest way to obtain a static generative model for multiplex networks is to generalize the existing methods for single-layer ones [1]. Fixing the degree sequence in each layer, one can use a configuration model to obtain a particular realization of the given set of connectivities. In [43] the authors have made the choice to add interlayer links arbitrarily. A different approach is to keep using a configuration model, but to specify the edges between the layers by means of a joint-degree distribution [44–46].

A similar method is to specify the degree sequences together with a probability matrix whose element  $(i, j)$  is the fraction of interlayer links between layers  $i$  and  $j$ . The actual link placement is still achieved via uniform random choice [47]. A generalization of this approach has been proposed in [48]; the authors impose the degree correlations within and between layers by means of a set of matrices that specify the fraction of edges between nodes of given degrees in given layers.

### Co-publications multilayer networks

By browsing the headers of scientific papers, we can represent and analyze three networks namely: authors’ networks, laboratories’ networks and institutions’ networks. We referred by header of paper, the description of the title of the paper, the names of authors and their affiliations (Fig. 2 is an example of paper’s header). The networks are defined as follows:



1. The network of authors. A node represents an author and, if an author  $i$  co-authored a paper with author  $j$ , the graph contains an undirected edge from  $i$  to  $j$ . If the paper is co-authored by  $k$  authors this generates a completely connected (sub)graph on  $k$  nodes.
2. The network of laboratories. A node represents a laboratory in which at least one author published a paper, an edge links two laboratories if it exists at least one paper co-authored by authors of these two laboratories.
3. The network of institutions. A node represents an institution of the authors who have published at least one paper and, an edge links two institutions if it exists at least one paper co-authored by authors from two laboratories each related to these institutions. The term institution is used to refer to a university, research center or a research institution.

Because of a reciprocal collaboration, these networks are undirected. They can be weighted by the number of publications between the two entities or not according to the goal that is given to the study. In this study, they are unweighted. Note that, for the type of nodes (author, laboratory or institution) considered, network construction is summarized in the subsequent creation of clique. The three generated networks are interdependent because of the affiliation relationship between their entities. So, in addition to the above description, we add new edges that represent affiliations' relations between authors and laboratories and affiliations' relations between laboratories and institutions.

We then say that these networks are multilayered (see Fig. 1) and are deduced from collaboration between authors. Indeed, the actors involved in co-publication are the authors. In these networks, we have two types of relationship: collaboration (at the same level) and affiliation (between two levels).

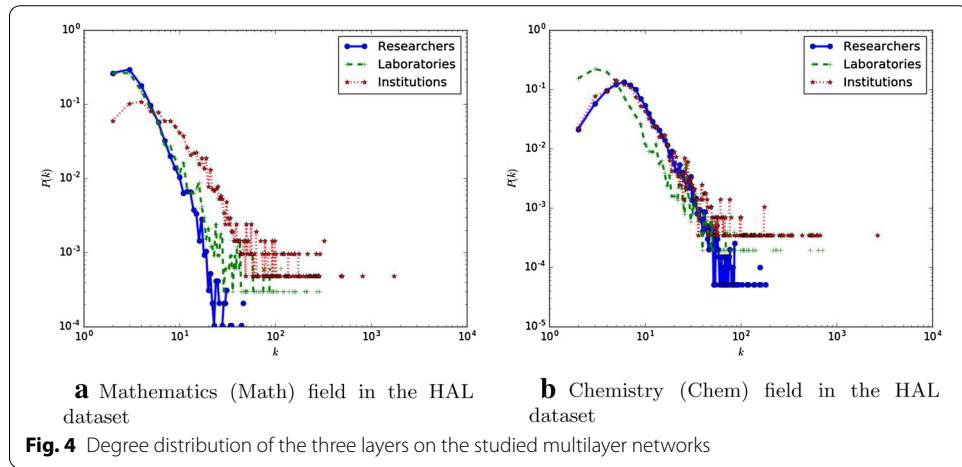
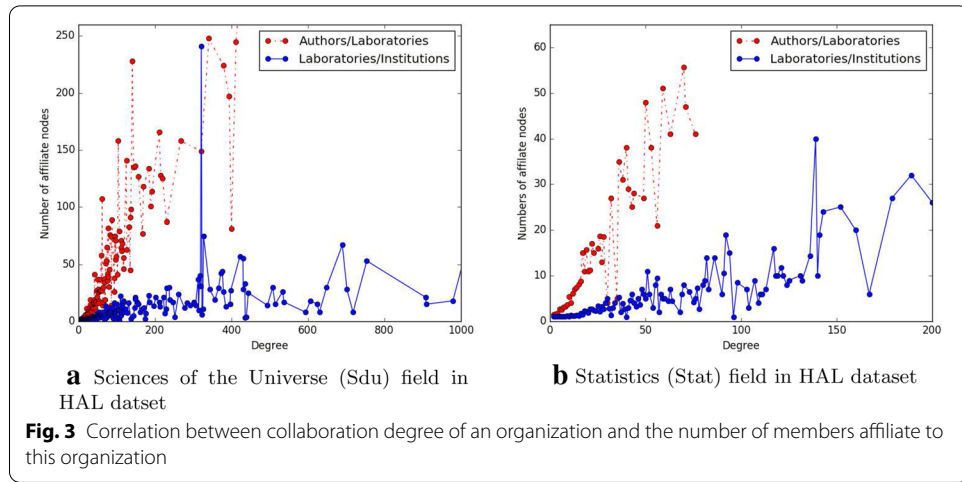
The studied networks can be formally represented by  $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ , where:  $\mathcal{G} = \{G_\alpha; \alpha \in \{1, 2, 3\}\}$  is a layers of  $\mathcal{M}$  with  $G_1$  is network of researchers,  $G_2$  is network of laboratories,  $G_3$  is network of institutions. Each  $G_\alpha; \alpha \in \{1, 2, 3\}$  is a collaboration network and  $\mathcal{C}$  is a set the affiliations edges.

We have built multilayer networks from the publications of the open archive HAL<sup>1</sup> between 2006 and 2016. HAL is an open archive where authors can deposit scholarly documents from all academic fields. The total number of papers used in this dataset is 70224 organized in eight research fields.

We have analyzed the average number of entities affiliated to an organization. Precisely, we looked the average number of researchers affiliated to a laboratory of a given degree and the number of laboratories affiliated with an institution of a given degree (see Fig. 3). Using linear regression model, we approximated the relation between the average number of entities affiliated to an organization of a given degree and the degree of organization. So the number of node at layer  $\alpha$ , affiliated to an organization at layer  $\alpha + 1$  with collaboration degree  $k$  can be defined as follows:

$$N_\alpha(k) = \rho_\alpha k + n_0^\alpha. \quad (1)$$

<sup>1</sup> <https://hal.archives-ouvertes.fr/>.



We have compared the degree distribution of the three layer (Fig. 4). We found that all the three layers have a degree distribution that follows power law. Looking the densities  $\delta_\alpha$  and the average distance  $l_\alpha$ ;  $\alpha \in \{1, 2, 3\}$  (Table 1), we found that, for all the fields of the HAL dataset:

$$\delta_3 \geq \delta_2 \geq \delta_1 \quad \text{and} \quad l_3 \leq l_2 \leq l_1. \quad (2)$$

As shown in Table 1, each of the three networks layers has a high clustering coefficient ( $C \approx 0.83$ ), a low average distance ( $l \approx 7$ ); they are small-world networks.

From all the observations made by analyzing the hierarchical network of the HAL dataset, we designed a network generation model that will reproduce the main properties of such type of networks.

## The hierarchical network generation model

### Collaboration and affiliation algorithms

Consider that each actor (elements involve in a collaborations) of our model is affiliated to at least one organization. An organization is also affiliated to at least one higher level



**Table 1** Properties of the layers: total number of nodes  $n$ ; total number of edges  $m$ ; mean degree  $\bar{d}$ ; mean node–node distance  $l$ ; exponent  $\gamma$  of degree distribution; clustering coefficient  $C$ ; Transitivity  $T$ , Density  $\delta$ 

	Network	$n$	$m$	$\bar{d}$	$l$	$\gamma$	$C$	$T$	$\delta$
Chem	Researchers	19794	80144	8.1	7.07	2.43	0.87	0.43	0.0004
	Laboratories	5064	14653	5.8	3.56	2.43	0.85	0.07	0.0011
	Institutions	4563	32045	14.058	2.10	2.44	0.9	0.03	0.003
Info	Researchers	26492	485275	36.6	10.2	2.44	0.92	0.74	0.0013
	Laboratories	9658	125832	26.0	4.48	2.43	0.87	0.84	0.0026
	Institutions	10463	163545	31.2	4.08	2.44	0.86	0.75	0.0029
Math	Researchers	9199	14236	3.09	13.8	2.40	0.80	0.54	0.0003
	laboratories	3164	8167	5.16	3.61	2.43	0.75	0.09	0.0016
	Institutions	3558	12299	6.91	3.49	2.43	0.75	0.11	0.0019
Phys	Researchers	19920	160915	16.1	9.32	2.43	0.89	0.80	0.0008
	Laboratories	4793	18199	7.59	4.19	2.43	0.82	0.34	0.0015
	Institutions	5161	22623	8.76	4.00	2.44	0.82	0.30	0.0016
Shs	Researchers	3955	6193	3.13	5.46	2.39	0.86	0.83	0.0007
	Laboratories	1729	3021	3.49	5.98	2.42	0.73	0.29	0.0020
	Institutions	1903	4142	4.35	5.12	2.43	0.73	0.23	0.0022
Sdu	Researchers	20873	98593	9.44	7.49	2.43	0.87	0.67	0.0004
	Laboratories	7781	29238	7.51	3.69	2.43	0.87	0.12	0.0009
	Institutions	8217	35408	8.61	3.63	2.44	0.86	0.12	0.0010
Sdv	Researchers	31557	154416	9.78	11.8	2.43	0.92	0.81	0.0003
	Laboratories	11010	46084	8.37	4.89	2.43	0.89	0.61	0.0007
	Institutions	11594	50942	8.78	4.69	2.43	0.88	0.52	0.0007
Stat	Researchers	2867	5876	4.10	9.41	2.40	0.87	0.57	0.0014
	Laboratories	1301	3229	4.96	4.20	2.41	0.80	0.25	0.0038
	Institutions	1541	5458	7.08	3.81	2.43	0.79	0.22	0.0045

Chem, Info, Math, Phys, Shs, Sdu, Sdv, Stat are academic fields of the HAL dataset

of organization. For simplification purpose, we suppose that, each actor or organization is affiliated to only one organization and the mobility of the actors (as for the authors in laboratories) is not considered. In this context, a node  $x^\alpha = (id, aff)$  at layer  $\alpha$  is represented by its ID  $id$  and the ID  $aff$  of his affiliation  $y^{\alpha+1}$  at a layer  $\alpha + 1$ . The actors belong to layer 0. The affiliations of actors are in layer 1. The affiliations of organizations at layer 1 is in layer 2 and so on.

To generate edges between node in the same layers, we propose a growth model for the multilayer collaboration network similar to that of Meleu et al. [33] (Algorithm 1). It is an iterative model that simulates at each step a collaboration between actor (node of layer 0) and creates relationships in networks. The collaboration at any layer  $\alpha \geq 1$  is deduced by the affiliations of the nodes at layer  $\alpha - 1$ .

**Algorithm 1:** Collaboration Model (CM)

---

```

1 for  $t = 1$  to  $N_a$  do
2    $n \leftarrow nb\_nodes(P)$ ;
3   for  $i = 1$  to  $n$  do
4      $p \leftarrow random()$ ;
5     if  $p \leq \lambda$  then
6       Select old node using preferential attachment;
7     else
8       create a new node ;
9   Create a clique between the  $n$  nodes

```

---

In Algorithm 1, the selection of old nodes is made according to preferential attachment; an old actor  $i$  of degree  $k_i$  (in-layer degree) is selected with probability proportional to  $P_i = k_i / \sum_j k_j$ . To create edges in other layers, we proceed recursively: from level 0, we select in layer 1 the affiliation of the nodes in the collaboration then, we create a clique with these nodes at level 1. We create edges in level 2 using the previous affiliation nodes and select their affiliation nodes at layer 2. . . This process is given in Algorithm 2.

Let us define the affiliation vector by the set  $\mathcal{V} = \{\lambda_0, \lambda_1, \dots, \lambda_{M-1}\}$ , where  $\lambda_\alpha, \alpha > 0$  is a probability of a new node at level  $\alpha - 1$  to be affiliated to an old node at level  $\alpha$  and  $\lambda_0$  is the proportion of old nodes by collaboration at level 0. We can observe that, if  $M = 1$ , this model (represented by Algorithm 2) is the same as to Meleu et al. 's model [33]. So the network at layer 0 has the properties described in model [33].

**Algorithm 2:** Multilayer Collaboration Model

---

```

1 for  $t = 1$  to  $N_a$  do
2    $n \leftarrow nb\_nodes(P)$ ;
3   for  $i = 1$  to  $n$  do
4      $p \leftarrow random()$ ;
5     if  $p \leq \lambda_0$  then
6       Select old actor using preferential attachment;
7     else
8       create a new actor and his affiliation using Algorithm 3;
9    $A_0 \leftarrow$  set of created and/or selected actors
10  for  $\alpha = 0$  to  $M - 1$  do
11    Create a clique between the nodes in  $A_0$  at level  $\alpha$ 
12     $A_0 \leftarrow$  affiliations of nodes in  $A_0$ 

```

---

When we create node at level 0, we decide, using affiliation vector to affiliate this node to old or new node at level 1. In the case of affiliation to a new node, we create node at level 1 and then decide (using the affiliation vector) to affiliate this node to an old or a new node at level 2. This is done recursively for the upper layers. We propose

in Algorithm 3, a process to create nodes and affiliate them to their organizations. The

---

**Algorithm 3:** Affiliation Model

---

```

1 Procedure Affiliation( $x^\alpha = [id, aff]$ ,  $\alpha$ ,  $\mathcal{V} = \{\lambda_0, \lambda_1, \dots, \lambda_{M-1}\}$ )
2   if  $\alpha = M$  then
3      $x^\alpha.aff \leftarrow 0$ ;
4     return
5   else
6      $p \leftarrow random()$ ;
7     if  $p \leq \lambda_{\alpha+1}$  then
8       Select old organization  $y^{\alpha+1}$  using preferential attachment;
9        $x^\alpha.aff \leftarrow y^{\alpha+1}.id$ ;
10    else
11      create a new organization  $y^{\alpha+1}$ ;
12       $x^\alpha.aff \leftarrow y^{\alpha+1}.id$ ;
13      Affiliation( $y^{\alpha+1}, \alpha + 1, \mathcal{V}$ )

```

---

node affiliation to organizations follows a preferential attachment.

A generated multilayer network is a pair  $\mathcal{M} = (\mathcal{G}, \mathcal{C})$ , where

- $\mathcal{G} = \{G_\alpha; \alpha \in \{0, 1, \dots, M-1\}\}$  is a family of collaborations graphs  $G_\alpha = (X_\alpha, E_\alpha)$ . This can be a collaboration networks of actors or collaboration network at each organization's level.
- $\mathcal{C} = \{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta; \alpha, \beta \in \{0, 1, \dots, M-1\}, \alpha \neq \beta\}$  is the set of affiliations between actors and organizations or between organizations and sub-organization  $G_\alpha$ .

**Properties of the generated networks**

Let  $\mathcal{M} = (\mathcal{G}, \mathcal{C})$  be a multilayer network generated by our model.

**Proposition 1** *The number of nodes in  $G_\alpha = (X_\alpha, E_\alpha)$  is:*

$$|X_\alpha| = t(1 - \lambda)(1 - \lambda_1) \dots (1 - \lambda_\alpha)\eta = t\eta \prod_{i=0}^{\alpha} (1 - \lambda_i) \quad \forall \alpha \in \{0, 1, \dots, M\}, \quad (3)$$

where  $t$  is the number of collaborations generated.

*Proof 1* We proceed by induction.

- 1 At layer 0, the network is the same as networks generated by Meleu et al. 's model [33], so:

$$|X_\alpha| = t(1 - \lambda_0)\eta.$$

- 2 Consider that, at layer  $\alpha$ ,  $0 \leq \alpha \leq M$ , we have

$$X_\alpha = t(1 - \lambda_0)(1 - \lambda_1) \dots (1 - \lambda_\alpha)\eta.$$

We will show that we have  $|X_{\alpha+1}| = t(1 - \lambda)(1 - \lambda_1) \dots (1 - \lambda_\alpha)\eta$  at layer  $\alpha + 1$ .

In fact, at each step

- $(1 - \lambda)\eta$  new nodes are added at layer 0.
- Those nodes generate  $(1 - \lambda_0)(1 - \lambda_1)\eta$  new nodes at layer 1 by affiliating new nodes of layer 0 with the new nodes in this layer 1. Since, each new node in layer 0 has the probability  $1 - \lambda_1$  (using Algorithm 3 and affiliation vector) to be affiliated to a new node in layer 1.
- By recurrence, at layer  $\alpha$  we have  $(1 - \lambda_0)(1 - \lambda_1) \dots (1 - \lambda_\alpha)\eta$  new nodes are added as affiliation of  $(1 - \lambda)(1 - \lambda_1) \dots (1 - \lambda_{\alpha-1})\eta$  new nodes of layer  $\alpha - 1$ . Those  $(1 - \lambda_0)(1 - \lambda_1) \dots (1 - \lambda_\alpha)\eta$  new nodes of layer  $\alpha$  will generate  $(1 - \lambda_0)(1 - \lambda_1) \dots (1 - \lambda_\alpha)(1 - \lambda_{\alpha+1})\eta$  new nodes in layer  $\alpha + 1$ .

Hence, for  $t$  steps, we find:

$$|X_{\alpha+1}| = t(1 - \lambda_0)(1 - \lambda_1) \dots (1 - \lambda_{\alpha+1})\eta.$$

□

**Proposition 2** *The number of edges  $|E_\alpha|$  in the network is:*

$$\begin{cases} \frac{t}{2}\eta \prod_{i=0}^{\alpha}(1 - \lambda_i)(1 + \eta \prod_{i=0}^{\alpha}(1 - \lambda_i)) \leq |E_\alpha| \leq \frac{t}{2}(\eta - 1)\eta & \text{if } \alpha \geq 1 \\ \frac{t}{2}(1 - \lambda)\eta[(1 - \lambda)\eta - 1] + t\lambda(1 - \lambda)\eta^2 \leq |E_\alpha| \leq \frac{t}{2}(\eta - 1)\eta & \text{if } \alpha = 0 \end{cases} \quad (4)$$

where  $t$  is the number of collaborations generated.

*Proof 2* While selecting or/and creating  $\eta$  actors by collaboration in layers 0, it is possible that all of them will be affiliated to different organizations in all the other layers  $\alpha \geq 1$ . Thus, the maximum number of edges created by a collaboration in each of the layer  $\alpha \geq 1$  is then:

$$\frac{1}{2}\eta(\eta - 1).$$

On other hand, at each step let us consider that all the old nodes of layer 0 involved in the clique creation are affiliated to the same organization i. e  $(1 - \lambda_0)\eta$  actors are affiliated to the same node  $x$  at layer 1. At this level(1), the number of new edges will be equal to the number of edges that link the organization  $x$  to all the new organizations added by the creation of a new node at level 0 (i. e affiliation of new actors). This number is:

$$\frac{1}{2}(1 - \lambda_0)(1 - \lambda_1)(1 + (1 - \lambda_0)(1 - \lambda_1)).$$

We have shown that, at layer  $\alpha$ ,  $0 \leq \alpha \leq M - 1$ , we have  $\prod_{i=0}^{\alpha}(1 - \lambda_i)$  new nodes. These new nodes generate  $\prod_{i=0}^{\alpha+1}(1 - \lambda_i)$  new nodes at layer  $\alpha + 1$  and we have assumed that, old nodes are affiliated to the same node at layer  $\alpha + 1$ , thus, the edges added at layer  $\alpha + 1$  is the edges of the clique of  $1 + \eta \prod_{i=0}^{\alpha}(1 - \lambda_i)$  nodes which is:

$$\frac{1}{2}\eta \prod_{i=0}^{\alpha}(1 - \lambda_i) \left( 1 + \eta \prod_{i=0}^{\alpha}(1 - \lambda_i) \right).$$

The result is deduced by considering  $t$  steps.

□

**Proposition 3** The average degree  $\bar{d}_\alpha$  of  $G_\alpha = (X_\alpha, E_\alpha)$  is:

$$\begin{cases} 1 + \eta \prod_{i=0}^{\alpha} (1 - \lambda_i) \leq \bar{d}_\alpha \leq \frac{\eta-1}{\prod_{j=0}^{\alpha} (1-\lambda_j)} & \text{if } \alpha \geq 1 \\ (1 + \lambda_0)\eta - 1 \leq \bar{d}_\alpha \leq \frac{\eta-1}{1-\lambda_0} & \text{if } \alpha = 0 \end{cases} \quad (5)$$

*Proof 3* By definition:

$$\bar{d} = \frac{2|E_\alpha|}{|X_\alpha|}.$$

The proposition is deduced from Eq(3) and Eq. (4).  $\square$

**Theorem 1** If the average degree at layer  $\alpha$  is  $\bar{d}_\alpha = (\eta - 1) / \left( \prod_{j=0}^{\alpha} (1 - \lambda_j) \right)$ , the degree distribution in layer  $\alpha$  of a generated multilayer network follows a power law of parameter  $\gamma_\alpha$  as:

$$\gamma_\alpha = \begin{cases} 1 + \frac{1}{\lambda_0} & \text{if } \alpha = 0 \\ 1 + \frac{1}{\lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1-\lambda_j)} & \text{if } \alpha \geq 1 \end{cases} \quad (6)$$

*Proof 4* For  $\alpha = 0$ , the result is shown in Ref. [33]. For  $\alpha > 0$ , according the hypothesis on the average degree: the probability that a new node at layer  $\alpha$  will connect to a node of the layer  $\alpha + 1$  is proportional to the degree of this last node in this  $\alpha + 1$  layer.

Since we assume Eq 1, the probability that a node  $x^{\alpha-1}$  is affiliated to an organization  $y^\alpha$  of degree  $k_y^\alpha$  is proportional to the degree of this node, i.e.:

$$P(x^{\alpha-1}, y_k^\alpha \text{ exists}) = \frac{N_\alpha(k)}{\sum_{t^\alpha} N_\alpha(k_t^\alpha)} \approx \frac{k}{\sum_{t^\alpha} k_t^\alpha}. \quad (7)$$

The variation of the nodes of degree  $k$  in layer  $\alpha$  is impacted by:

- the selection of nodes in layer  $\alpha - 1$  affiliated to node of layer  $\alpha$  having degree  $k$ .
- the selection of old nodes of degree  $k$  to affiliate new nodes of layer  $\alpha - 1$  using affiliation vector.

It follows that, the number of nodes of degree  $k$  at step  $t$  in layer  $\alpha$  that gain an edge when the algorithm creates a new collaboration is:

$$N_k = \eta \left( \lambda + \sum_{i=1}^{\alpha-1} \lambda_i \prod_{j<i} (1 - \lambda_j) \right) P(x^{\alpha-1} \in y_k^\alpha) + \frac{\eta \lambda_\alpha \prod_{i<\alpha} (1 - \lambda_i)}{\sum_x x P_x} k P_k.$$

Using Eq. 7 we obtain

$$\frac{\eta \left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)}{\sum_x x P_x} k P_k.$$

If we denote by  $p_{k,t}$  the value of  $p_k$  when the network has  $n_t$  vertices, then the variation in  $n_t p_k$  per  $\prod_{i=0}^{\alpha} (1 - \lambda_i) \eta$  vertices added is:

$$\begin{aligned} & \left( n_t + \eta \prod_{i=0}^{\alpha} (1 - \lambda_i) \right) p_{k,t+1} - n_t p_{k,t} \\ &= \frac{\eta \left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)}{\sum_x x P_x} ((k - \eta + 1) p_{k-\eta+1,t} - k p_{k,t}). \end{aligned} \quad (8)$$

Looking for stationary solutions  $p_{k,t+1} = p_{k,t} = p_k$  as, the variation of the number of nodes of degree  $k$  at layer  $\alpha$  is then:

$$\eta \prod_{j=0}^{\alpha} (1 - \lambda_j) P_k = \frac{\eta \left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)}{\sum_x x P_x} [(k - \eta + 1) P_{k-\eta+1} - k P_k] \quad (9)$$

with:

$$\sum_x x P_x = \bar{d}_{\alpha} = \frac{\eta - 1}{\prod_{j=0}^{\alpha} (1 - \lambda_j)}.$$

By simplifying in Eq. (9) we obtain:

$$\begin{aligned} p_k &= \frac{k - \eta + 1}{k + \frac{\eta - 1}{\left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)}} P_{k-\eta+1} \\ &= \frac{\frac{k}{\eta - 1} - 1}{\frac{k}{\eta - 1} + \frac{1}{\left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)}} \cdots \frac{1}{2 + \frac{1}{\left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)}} P_{\eta-1} \\ &= c B \left( \frac{k}{\eta - 1}, 1 + \frac{1}{\left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)} \right), \end{aligned}$$

where  $c = P_{\eta-1}$  et  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$  is Legendre's beta-function, which goes asymptotically as  $a^{-b}$  for large  $a$  and fixed  $b$ , and hence

$$p_k \approx \left( \frac{k}{\eta - 1} \right)^{- \left( 1 + \frac{1}{\lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j)} \right)}.$$

□



**Theorem 2** *If the average degree is  $1 + \eta \prod_{i=0}^{\alpha} (1 - \lambda_i)$ , the degree distribution in layer  $\alpha > 0$  of a generated multilayer network follows a power law of parameter  $\gamma_{\alpha}$  as:*

$$\gamma_{\alpha} = 1 + \frac{\frac{1}{\eta} + \prod_{j=0}^{\alpha} (1 - \lambda_j)}{\lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j)}. \quad (10)$$

*Proof 5* Let  $\theta = \prod_{j=0}^{\alpha} (1 - \lambda_j)$ ; similarly to the previous case, and looking for stationary solutions  $p_{k,t+1} = p_{k,t} = p_k$  as, the variation of the number of nodes of degree  $k$  at layer  $\alpha$  is:

$$\eta \theta P_k = \frac{\eta \left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)}{\sum_x x P_x} [(k - \eta \theta) P_{k-\eta \theta} - k P_k] \quad (11)$$

with:

$$\sum_x x P_x = \bar{d}_{\alpha} = 1 + \eta \prod_{j=0}^{\alpha} (1 - \lambda_j).$$

Thus:

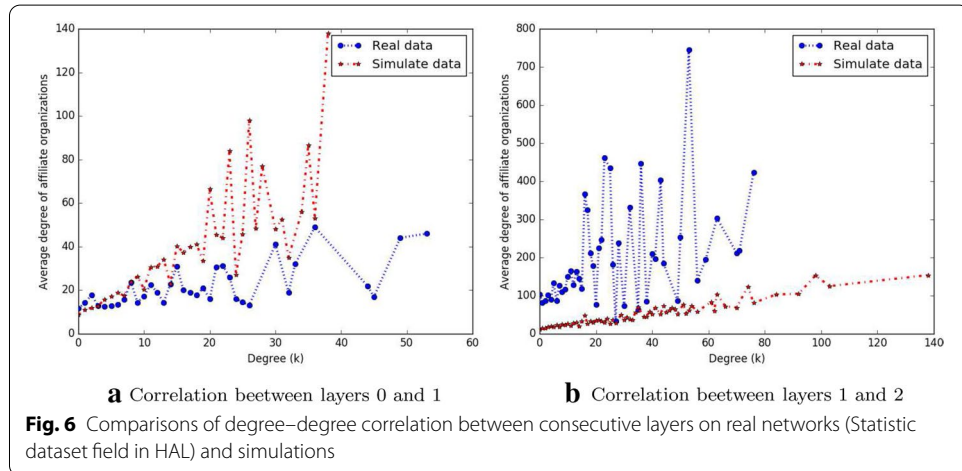
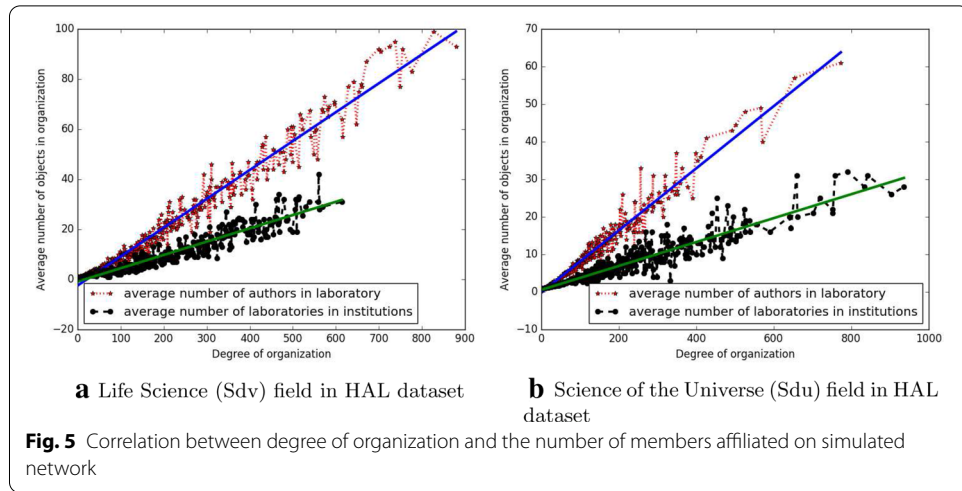
$$\eta \theta P_k = \frac{\eta \left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right)}{1 + \eta \theta} [(k - \eta \theta) P_{k-\eta \theta} - k P_k], \quad (12)$$

$$\theta (1 + \eta \theta) P_k = \left( \lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j) \right) ((k - \eta \theta) P_{k-\eta \theta} - k P_k). \quad (13)$$

By simplifying in Eq. (13), we obtain

$$\begin{aligned} p_k &= \frac{k - \eta \theta}{k + \frac{\theta(1+\eta\theta)}{\left(\lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j)\right)}} P_{k-\eta\theta} \\ &= \frac{\frac{k}{\eta\theta} - 1}{\frac{k}{\eta\theta} + \frac{1+\eta\theta}{\eta \left(\lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j)\right)}} \cdots \frac{1}{2 + \frac{1+\eta\theta}{\eta \left(\lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j)\right)}} P_{\eta\theta} \\ &= cB \left( \frac{k}{\eta \prod_{j=0}^{\alpha} (1 - \lambda_j)}, 1 + \frac{1 + \eta \prod_{j=0}^{\alpha} (1 - \lambda_j)}{\eta \left(\lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j<i} (1 - \lambda_j)\right)} \right), \end{aligned}$$

where  $c = P_{\eta \prod_{j=0}^{\alpha} (1 - \lambda_j)}$  et  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$  is Legendre's beta-function, which goes asymptotically as  $a^{-b}$  for large  $a$  and fixed  $b$ , and hence:

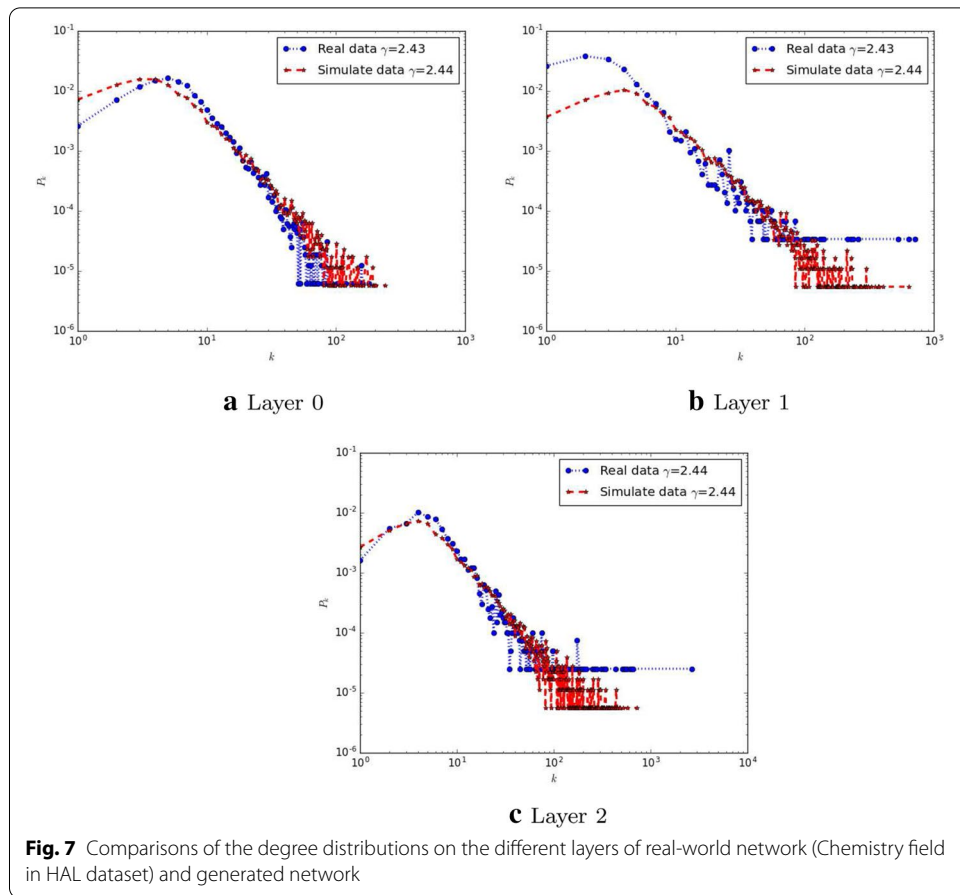


$$p_k \approx \left( \frac{k}{\eta \prod_{j=0}^{\alpha} (1 - \lambda_j)} \right) - \left( 1 + \frac{\frac{1}{\eta} + \prod_{j=0}^{\alpha} (1 - \lambda_j)}{\lambda_0 + \sum_{i=1}^{\alpha} \lambda_i \prod_{j < i} (1 - \lambda_j)} \right).$$

□

## Simulations

We have implemented our model using a custom-written Java program. The outputs of this program were used by a custom program written in python and based

**Table 2** Parameters of the model

Designation	Description
$M$	Number of layers
$N_a$	Number of collaborations to generate
$P$	The distribution of the number of actors per collaboration $P_i$ = probability to have $i$ actors in a collaboration
$\mathcal{V} = \{\lambda_0, \lambda_1, \dots, \lambda_M\}$	Affiliation vector

on the libraries: NetworkX<sup>2</sup>, Pymnet<sup>3</sup> and powerlaw [49]; to make most of the measurement on the generated networks. To generate the networks, we extracted parameters such as number of collaborations to generate, the affiliation vector and the distribution of numbers of actors by collaboration from the different fields of HAL dataset.

<sup>2</sup> <https://networkx.github.io/>.

<sup>3</sup> [http://people.maths.ox.ac.uk/kivela/mln\\_library/](http://people.maths.ox.ac.uk/kivela/mln_library/).

**Table 3 Properties of the simulated networks layers: total number of nodes  $n$ ; total number of edges  $m$ ; mean degree  $\bar{d}$ ; mean node–node distance  $l$ ; exponent  $\gamma$  of degree distribution; clustering coefficient  $C$ ; Transitivity  $T$ , Density  $\delta$** 

	Network	$n$	$m$	$\bar{d}$	$l$	$\gamma$	$C$	$T$	$\delta$
Shs	Researchers	3220	6409	3.98	4.53	2.41	0.65	0.19	0.0012
	Laboratories	2591	6290	4.85	3.95	2.42	0.62	0.13	0.0018
	Institutions	2291	6175	5.39	3.76	2.43	0.60	0.12	0.0023
Stat	Researchers	3310	6879	4.15	5.53	2.40	0.80	0.34	0.0012
	Laboratories	2066	6765	6.55	3.77	2.43	0.68	0.14	0.0031
	Institutions	1363	6534	9.59	3.21	2.43	0.63	0.14	0.0070
Math	Researchers	9137	15813	3.46	5.67	2.42	0.71	0.20	0.0003
	Laboratories	7134	15732	4.41	4.61	2.43	0.65	0.11	0.0006
	Institutions	5710	15607	5.46	4.07	2.43	0.62	0.08	0.0009
Sdv	Researchers	34698	164777	9.49	5.45	2.43	0.90	0.57	0.0002
	Laboratories	20154	163133	16.1	3.52	2.44	0.79	0.16	0.0008
	Institutions	13656	157545	23.0	3.07	2.44	0.75	0.12	0.0016
Phys	Researchers	29681	527031	35.5	3.32	2.44	0.89	0.32	0.0011
	Laboratories	22775	512941	45.0	3.05	2.44	0.86	0.22	0.0019
	Institutions	17647	482890	54.7	2.87	2.44	0.85	0.20	0.0031
Sdu	Researchers	21878	117614	10.7	4.06	2.43	0.83	0.26	0.0004
	Laboratories	7069	98700	27.9	2.86	2.44	0.75	0.18	0.0039
	Institutions	2430	62543	51.4	2.52	4.43	0.80	0.39	0.0211
Chem	Researchers	20603	92155	8.94	4.36	2.43	0.80	0.22	0.0004
	Laboratories	15326	91388	11.9	3.65	2.44	0.76	0.11	0.0007
	Institutions	11321	89263	15.7	3.26	2.44	0.73	0.10	0.0013

Chem, Info, Math, Phys, Shs, Sdu, Sdv, Stat are academic fields of the HAL dataset

The first behavior that we wanted to observe on the simulated networks is the correlation between the in-layer degree of an organization and the number of members affiliated. Using linear regressions, we compared this correlation between real networks (Fig. 3) and the generated one (Fig. 5) and, it appears that, their values are close.

Figure 6 shows the comparison of degree–degree correlation between consecutive layers. Layer 0 is the collaboration network of researchers, layer 1 is the collaboration network of laboratories and the layers 2 the collaboration network of institutions. We can see that on real networks and simulations the curves can correctly be approximated using linear regression. The positive slopes of the obtained regression line mean that researchers, with high number of collaborations are affiliated to organizations with strong capacity of cooperation. This is easily explained on the simulations by the fact that, since a node is affiliated to a single organization, each time this node participates in collaborations, it induces participation in the collaborations of its affiliations organizations.

In Fig. 7, we have depicted the degree distributions of the different layers. We found that the simulations reproduce the power law distributions observed on the three layers of HAL dataset. Indeed, the exponent ( $\gamma$ ) of the power law degree-distribution in Tables 1, 2 and 3 are close. From Table 3 we observe that, the simulated layers networks have very high clustering coefficients ( $C \approx 0.80$ ) and high transivities ( $T \approx 0.2$ ). This behavior contributes to create high density and high average degree

in the affiliation layers in comparison with the current layer. So the density grows inversely than average distance. We can conclude that, the simulated networks layers are the small-world network.

## Conclusion

In this article, we have shown that headers of scientific papers can be used to build co-publication networks that are multilayered networks of entities involved in research namely: authors, laboratories, and institutions. Indeed, in addition to the collaborative relationships that exist between entities of each type, there are affiliation relationships of authors to laboratories and laboratories to institutions. We have analyzed the properties of such networks built from data extracted from the HAL platform which is a free archive of scientific publications.

Following the observations made on the properties of co-publications multilayer networks, we generalized these networks to a system of actors and organizations such that an actor is affiliated to an organization and each organization is affiliated with a higher level organization. We then said that the graphs are hierarchical and are deduced from the collaboration between the actors. The actors collaborate together and the relationships in the different layers are deduced from these collaboration and their affiliation relationships. We proposed an algorithmic model to build graphs presenting such properties. It is an iterative model that builds a collaboration clique and related affiliations at each step. We showed that the degree distribution in different layers follows a power law and the simulations carried out showed that the studied properties of the generated layers are close to those of the the real-world network built from HAL dataset.

In the future, we are planning to explore the structure and dynamics of communities in such co-publication networks. Indeed, the high clustering coefficient and high transitivity in these graphs suggest the existence of many communities. Before doing this, we will verify the robustness of our model by using other scientific publications archives and perform a more accurate evaluation of the gaps between the values of the properties of the generated networks and those of real-world networks.

## Acknowledgements

We would like to sincerely thank the IDASCO members who provided constructive comments.

## Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Authors' contributions

For this article, GRM made: conception or design of the models, data collection, data analysis and interpretation, drafting the article, critical revision of the article, final approval of the version to be published. PYM made: conception or design of the models, critical revision of the article, final approval of the version to be published. Both authors read and approved the final manuscript.

## Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

## Data availability statement

The datasets analyzed during the current study were extracted from The open archive HAL: <https://hal.archives-ouvertes.fr/>.

## Competing interests

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup> Département de Mathématiques et Informatique, Faculté des Sciences, Université de Ngaoundéré, B. P. 454 Ngaoundéré, Cameroun. <sup>2</sup> Département d'Informatique, Faculté des Sciences, Université de Yaoundé 1, B. P. 812 Yaoundé, Cameroun. <sup>3</sup> IRD, UMIMISCO, 93143 Bondy, France.

Received: 2 January 2020 Accepted: 21 January 2021

Published online: 09 February 2021

**References**

1. Boccaletti S, Bianconi G, Criado R, Del Genio CI, Gómez-Gardeñes J, Romance M, Sendiña-Nadal I, Wang Z, Zanin M. The structure and dynamics of multilayer networks. *Phys Rep*. 2014;544(1):1–122.
2. Lattanzi S, Sivakumar D. Affiliation networks. In: *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, ACM; 2009. p. 427–434.
3. Zitnik M, Leskovec J. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics*. 2017;33(14):190–8.
4. Gallotti R, Barthelemy M. Anatomy and efficiency of urban multimodal mobility. *Sci Rep*. 2014;4:6911.
5. Cardillo A, Zanin M, Gómez-Gardenes J, Romance M, del Amo AJG, Boccaletti S. Modeling the multi-layer nature of the european air transport network: Resilience and passengers re-scheduling under random failures. *Eur Phys J Spec Top*. 2013;215(1):23–33.
6. Battiston F, Nicosia V, Latora V. Structural measures for multiplex networks. *Phys Rev E*. 2014;89(3):032804.
7. Hristova D, Noulas A, Brown C, Musolesi M, Mascolo C. A multilayer approach to multiplexity and link prediction in online geo-social networks. *EPJ Data Sci*. 2016;5(1):24.
8. Kurucz M, Benczur A, Csalogány K, Lukács L. Spectral clustering in telephone call graphs. In: *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, ACM; 2007. p. 82–91.
9. Berge C. *Hypergraphs: Combinatorics of Finite Sets*, vol. 45. New York: Elsevier; 1984.
10. Kamga V, Tchuente M, Viennet E. Prédiction de liens dans les graphes bipartites avec attributs. In: *AAFD*; 2012. p. 57–70.
11. Ngonmang B, Viennet E, Tchuente M, Kamga V. Community analysis and link prediction in dynamic social networks. In: *Computing in Research and Development in Africa*, Springer; 2015. p. 83–101.
12. De Domenico M, Solé-Ribalta A, Cozzo E, Kivelä M, Moreno Y, Porter MA, Gómez S, Arenas A. Mathematical formulation of multilayer networks. *Phys Rev X*. 2013;3(4):041022.
13. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. *J Compl Netw*. 2014;2(3):203–71.
14. Solá L, Romance M, Criado R, Flores J, García del Amo A, Boccaletti S. Eigenvector centrality of nodes in multiplex networks. *Chaos*. 2013;23(3):033131.
15. Holme P, Saramäki J. Temporal networks. *Phys Rep*. 2012;519(3):97–125.
16. Donges J, Schultz H, Marwan N, Zou Y, Kurths J. Investigating the topology of interacting networks. *Eur Phys J B*. 2011;4(84):635–51.
17. Berlingerio M, Coscia M, Giannotti F, Monreale A, Pedreschi D. Multidimensional networks: foundations of structural analysis. *World Wide Web*. 2013;16(5–6):567–93.
18. Coscia M, Rossetti G, Pennacchioli D, Ceccarelli D, Giannotti F. “you know because i know”: A multidimensional network approach to human resources problem. In: *Advances in Social Networks Analysis and Mining (ASONAM)*, 2013 IEEE/ACM International Conference On, 2013; IEEE. p. 434–441.
19. Gao J, Buldyrev SV, Stanley HE, Havlin S. Networks formed from interdependent networks. *Nat Phys*. 2012;8(1):40–8.
20. Criado R, Flores J, García del Amo A, Gómez-Gardeñes J, Romance M. A mathematical model for networks with structures in the mesoscale. *Int J Comput Math*. 2012;89(3):291–309.
21. Berlingerio M, Coscia M, Giannotti F, Monreale A, Pedreschi D. Foundations of multidimensional network analysis. In: *Advances in Social Networks Analysis and Mining (ASONAM)*, 2011 International Conference On, IEEE; 2011. p. 485–489.
22. Watts DJ. The dynamics of networks between order and randomness. *Small Worlds* 1999.
23. Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *Nature*. 1998;393(6684):440–2.
24. Milgram S. The small world problem. *Psychol Today*. 1967;2(1):60–7.
25. De Stefano D, Giordano G, Vitale MP. Issues in the analysis of co-authorship networks. *Qual Quant*. 2011;45(5):1091–107.
26. De Stefano D, Fuccella V, Vitale MP, Zaccarin S. The use of different data sources in the analysis of co-authorship networks and scientific performance. *Soc Netw*. 2013;35(3):370–81.
27. Mali F, Kronegger L, Doreian P, Ferligoj A. Dynamic scientific co-authorship networks. In: *Models of Science Dynamics*, Springer; 2012. p. 195–232.
28. Erdos P, Rényi A. On the evolution of random graphs. *Publ Math Inst Hung Acad Sci*. 1960;5(1):17–60.
29. Molloy M, Reed B. The size of the giant component of a random graph with a given degree sequence. *Combin Prob Comput*. 1998;7(03):295–305.
30. Guillaume J-L, Latapy M. Bipartite graphs as models of complex networks. *Physica A*. 2006;371(2):795–813.
31. Price DJS. A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inform Sci*. 1976;29:3:1.
32. Barabási A-L, Jeong H, Neda Z, Ravasz E, Schubert A, Vicsek T. Evolution of the social network of scientific collaborations. *Physica A*. 2002;311(3):590–614.
33. Meleu GR, Melatagia Yonta P. Growth model for collaboration networks. *Revue Africaine de la Recherche en Informatique et Mathématiques Appliqués*. 2017;24:1–21.



34. Newman ME. The structure and function of complex networks. *SIAM Rev.* 2003;45(2):167–256.
35. Nicosia V, Bianconi G, Latora V, Barthélemy M. Growing multiplex networks. *Phys Rev Lett.* 2013;111(5):058701.
36. Kim JY, Goh K-I. Coevolution and correlated multiplexity in multiplex networks. *Phys Rev Lett.* 2013;111(5):058702.
37. Magnani M, Rossi L. The ml-model for multi-layer social networks. In: *Advances in Social Networks Analysis and Mining (ASONAM)*, 2011 International Conference On, 2011; IEEE. p. 5–12
38. Bianconi G. Statistical mechanics of multiplex networks: Entropy and overlap. *Phys Rev E.* 2013;87(6):062806.
39. Pattison P, Wasserman S. Logit models and logistic regressions for social networks. *Br J Math Stat Psychol.* 1999;52(2):169–93.
40. Wang P, Robins G, Pattison P, Lazega E. Exponential random graph models for multilevel networks. *Soc Netw.* 2013;35(1):96–115.
41. Zhang P-P, Chen K, He Y, Zhou T, Su B-B, Jin Y, Chang H, Zhou Y-P, Sun L-C, Wang B-H, et al. Model and empirical study on some collaboration networks. *Physica A.* 2006;360(2):599–616.
42. Nicosia V, Latora V. Measuring and modeling correlations in multiplex networks. *Phys Rev E.* 2015;92(3):032805.
43. Min B, Do Yi S, Lee K-M, Goh K-I. Network robustness of multiplex networks with interlayer degree correlations. *Phys Rev E.* 2014;89(4):042811.
44. Lee K-M, Kim JY, Cho W-K, Goh K-I, Kim I. Correlated multiplexity and connectivity of multiplex random networks. *New J Phys.* 2012;14(3):033027.
45. Funk S, Jansen VA. Interacting epidemics on overlay networks. *Phys Rev E.* 2010;81(3):036118.
46. Marceau V, Noël P-A, Hébert-Dufresne L, Allard A, Dubé LJ. Modeling the dynamical interaction between epidemics on overlay networks. *Phys Rev E.* 2011;84(2):026105.
47. Söderberg B. Properties of random graphs with hidden color. *Phys Rev E.* 2003;68(2):026107.
48. Melnik S, Porter MA, Mucha PJ, Gleeson JP. Dynamics on modular networks with heterogeneous correlations. *Chaos.* 2014;24(2):023106.
49. Alstott J, Bullmore E, Plenz D. Powerlaw: a python package for analysis of heavy-tailed distributions. *PLoS ONE.* 2014;9(1):85777.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)