



# Weighted Multiplex Networks

Giulia Menichetti<sup>1</sup>, Daniel Remondini<sup>1</sup>, Pietro Panzarasa<sup>2</sup>, Raúl J. Mondragón<sup>3</sup>, Ginestra Bianconi<sup>4\*</sup>

**1** Department of Physics and Astronomy and INFN Sez. Bologna, Bologna University, Bologna, Italy, **2** School of Business and Management, Queen Mary University of London, London, United Kingdom, **3** School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom, **4** School of Mathematical Sciences, Queen Mary University of London, London, United Kingdom

## Abstract

One of the most important challenges in network science is to quantify the information encoded in complex network structures. Disentangling randomness from organizational principles is even more demanding when networks have a multiplex nature. Multiplex networks are multilayer systems of  $N$  nodes that can be linked in multiple interacting and co-evolving layers. In these networks, relevant information might not be captured if the single layers were analyzed separately. Here we demonstrate that such partial analysis of layers fails to capture significant correlations between weights and topology of complex multiplex networks. To this end, we study two weighted multiplex co-authorship and citation networks involving the authors included in the American Physical Society. We show that in these networks weights are strongly correlated with multiplex structure, and provide empirical evidence in favor of the advantage of studying weighted measures of multiplex networks, such as multistrength and the inverse multiparticipation ratio. Finally, we introduce a theoretical framework based on the entropy of multiplex ensembles to quantify the information stored in multiplex networks that would remain undetected if the single layers were analyzed in isolation.

**Citation:** Menichetti G, Remondini D, Panzarasa P, Mondragón RJ, Bianconi G (2014) Weighted Multiplex Networks. PLoS ONE 9(6): e97857. doi:10.1371/journal.pone.0097857

**Editor:** Yamir Moreno, University of Zaragoza, Spain

**Received:** February 19, 2014; **Accepted:** April 25, 2014; **Published:** June 6, 2014

**Copyright:** © 2014 Menichetti et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** DR and GM acknowledge support by the Italian Ministry of Education and Research through the Flagship (PB05) InterOmics and the European Methods for Integrated analysis of multiple Omics datasets (MIMOmics) (305280) projects. The funders had no role in study design and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: ginestra.bianconi@gmail.com

## Introduction

Network theory investigates the global topology and structural patterns of the interactions among the constituent elements of a number of complex systems including social groups, infrastructure and technological systems, the brain and biological networks [1–4]. Over the last fifteen years, a large body of literature has attempted to disentangle noise and stochasticity from non-random patterns and mechanisms, in an attempt to gain a better understanding of how these systems function and evolve. More recently, further advances in the study of complex systems have been spurred by the upsurge of interest in multiplex networks in which pairs of interacting elements are represented as nodes connected through multiple types of links, at multiple points in time, or at multiple scales of resolution [5]. More specifically, a multiplex network is a set of  $N$  nodes interacting in  $M$  layers, each reflecting a distinct type (or time or resolution) of interaction linking the same pair of nodes. Examples of multiplex networks include: social networks, where the same individuals can be connected through different types of social ties originating from friendship, collaboration, or family relationships [6]; air transportation networks, where different airports can be connected through flights of different companies [7]; and the brain, where different regions can be seen as connected by the functional and structural neural networks [8].

Most of the studies so far conducted on multiplex networks have been concerned with the empirical analysis of a wide range of systems [6,7,9,10], the modeling of their underlying structures [11–13], and the description of new critical phenomena and

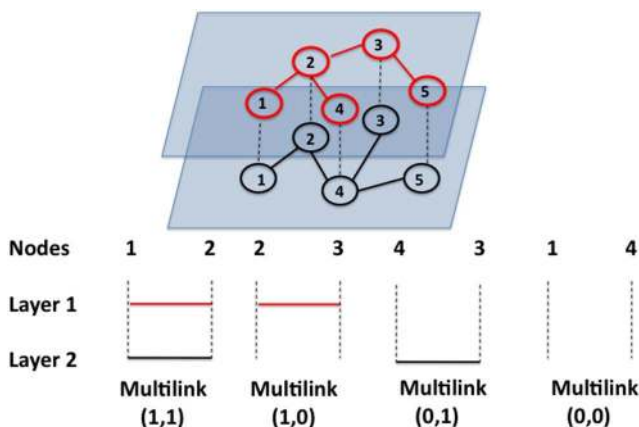
processes occurring on them [14–17]. Despite the growing interest in multiplex networks, a fundamental question still remains largely unanswered: What is the advantage of a full-fledged analysis of complex systems that takes all their interacting layers into account, over more traditional studies that represent such systems as single networks with only one layer? To answer this question, one should demonstrate that novel and relevant information can be uncovered only by taking the multiplex nature of complex systems directly into account, and would instead remain undetected if individual layers were analyzed in isolation. In this paper, an attempt is made to offer a possible solution to this problem within the context of weighted multiplex networks.

Just as with single networks, links between nodes may have a different weight, reflecting their intensity, capacity, duration, intimacy or exchange of services [18]. The role played by the weights in the functioning of many networks, and especially the relative benefits of weak and strong ties in social networks, have been the subject of a longstanding debate [18–20]. Moreover, it has been shown that, in single networks, the weights can be distributed in a heterogeneous way, as a result of the non-trivial effects that the structural properties of the networks have on them [21–24]. In particular, correlations between weights and structural properties of single networks can be uncovered by the analysis of strength-degree correlations [21] and by the distribution of the weights of the links incident upon the same node [23]. To characterize weighted networks, it is common practice to measure the following quantities: i) the average strength of nodes of degree  $k$ , i.e.  $s = s(k)$ , describing how weights are distributed in the

network; and ii) the average inverse participation ratio of the weights of the links incident upon nodes of degree  $k$ , i.e.  $Y = Y(k)$ , describing how weights are distributed across the links incident upon nodes of degree  $k$ . Here we show that these two quantities do not capture the full breadth of the information encoded in multiplex networks. Indeed, a full-fledged analysis of the properties of multiplex networks is needed that takes the multiple interacting and co-evolving layers simultaneously into account.

For a multiplex network, a *multilink*  $\vec{m} = (m_1, m_2, \dots, m_M)$  between nodes  $i$  and  $j$  indicates the set of all links connecting these nodes in the different layers [25]. In particular, if  $m_\alpha = 1$ , there is a link between nodes  $i$  and  $j$  in layer  $\alpha$ , whereas if  $m_\alpha = 0$  nodes  $i$  and  $j$  are not connected in layer  $\alpha$ . Multilink  $\vec{m} = \vec{0}$  between two nodes refers to the case in which no link exists between the two nodes in all layers of the multiplex network. Thus, multilinks indicate the most straightforward type of correlation between layers, and provide a simple generalization of the notion of overlap. In fact, if nodes  $i$  and  $j$  are connected by a multilink  $\vec{m}$ , with  $m_\alpha = m_{\alpha'} = 1$ , it follows that there is an overlap of links between  $i$  and  $j$  in layers  $\alpha$  and  $\alpha'$ . Figure 1 shows a multiplex network with  $M = 2$  layers and  $N = 5$  nodes with different types of multilinks.

Here we will define two new measures, *multistrength* and the *inverse multiparticipation ratio*, which are, respectively, the sum of the weights of a certain type of multilink incident upon a single node and a way for characterizing the heterogeneity of the weights of multilink  $\vec{m}$  incident upon a single node. To provide empirical evidence that weighted properties of multilinks are fundamental for properly assessing weighted multiplex networks, we focus on the networks of the authors of papers published in the journals of the American Physical Society (APS), and analyze the scientific collaboration network and the citation network connecting the same authors. These networks are intrinsically weighted since any two scientists can co-author more than one paper and can cite each other's work several times. A large number of studies have analyzed similar bibliometric datasets drawing upon network theory [26–30]. Unlike these studies, here we investigate the APS bibliometric dataset using the framework of multiplex networks that allows us to explore novel properties of the collaboration and citation networks. In particular, we show that multistrength and the inverse multiparticipation ratio enable new relevant information to be extracted from the APS dataset and that this information extends beyond what is encoded in the strength and inverse participation



**Figure 1. Example of all possible multilinks in a multiplex network with  $M = 2$  layers and  $N = 5$  nodes.** Nodes  $i$  and  $j$  are linked by one multilink  $\vec{m} = (m_\alpha, m_{\alpha'})$ . doi:10.1371/journal.pone.0097857.g001

ratio of single layers. Finally, based on the entropy of multiplex ensembles, we propose an indicator  $\Xi$  to evaluate the additional amount of information that can be extracted from the weighted properties of multilinks in multiplex networks over the information encoded in the properties of their individual layers analyzed separately.

## Weighted Multiplex Networks

### 2.1 Definition

A weighted multiplex network is a set of  $M$  weighted networks  $G_\alpha = (V, E_\alpha)$ , with  $\alpha = 1, \dots, M$ . The set of nodes  $V$  is the same for each layer and has cardinality  $|V| = N$ , whereas the set of links  $E_\alpha$  depends on the layer  $\alpha$ . A multiplex network is represented formally as  $\vec{G} = (G_1, G_2, \dots, G_\alpha, \dots, G_M)$ . Each network  $G_\alpha$  is fully described by the adjacency matrix  $\mathbf{a}^\alpha$  with elements  $a_{ij}^\alpha$ , where  $a_{ij}^\alpha = w_{ij}^\alpha > 0$  if there is a link with weight  $w_{ij}^\alpha$  between nodes  $i$  and  $j$  in layer  $\alpha$ , and  $a_{ij}^\alpha = 0$  otherwise. From now on, in order to simplify the formalization of weighted multiplex networks, we will assume that the weight of the link between any pair of nodes  $i$  and  $j$ ,  $a_{ij}^\alpha = w_{ij}^\alpha$ , can only take integer values. This does not represent a major limitation because in a large number of weighted multiplex networks the weights of the links can be seen as multiples of a minimal weight.

### 2.2 Structural Properties of Individual Layers

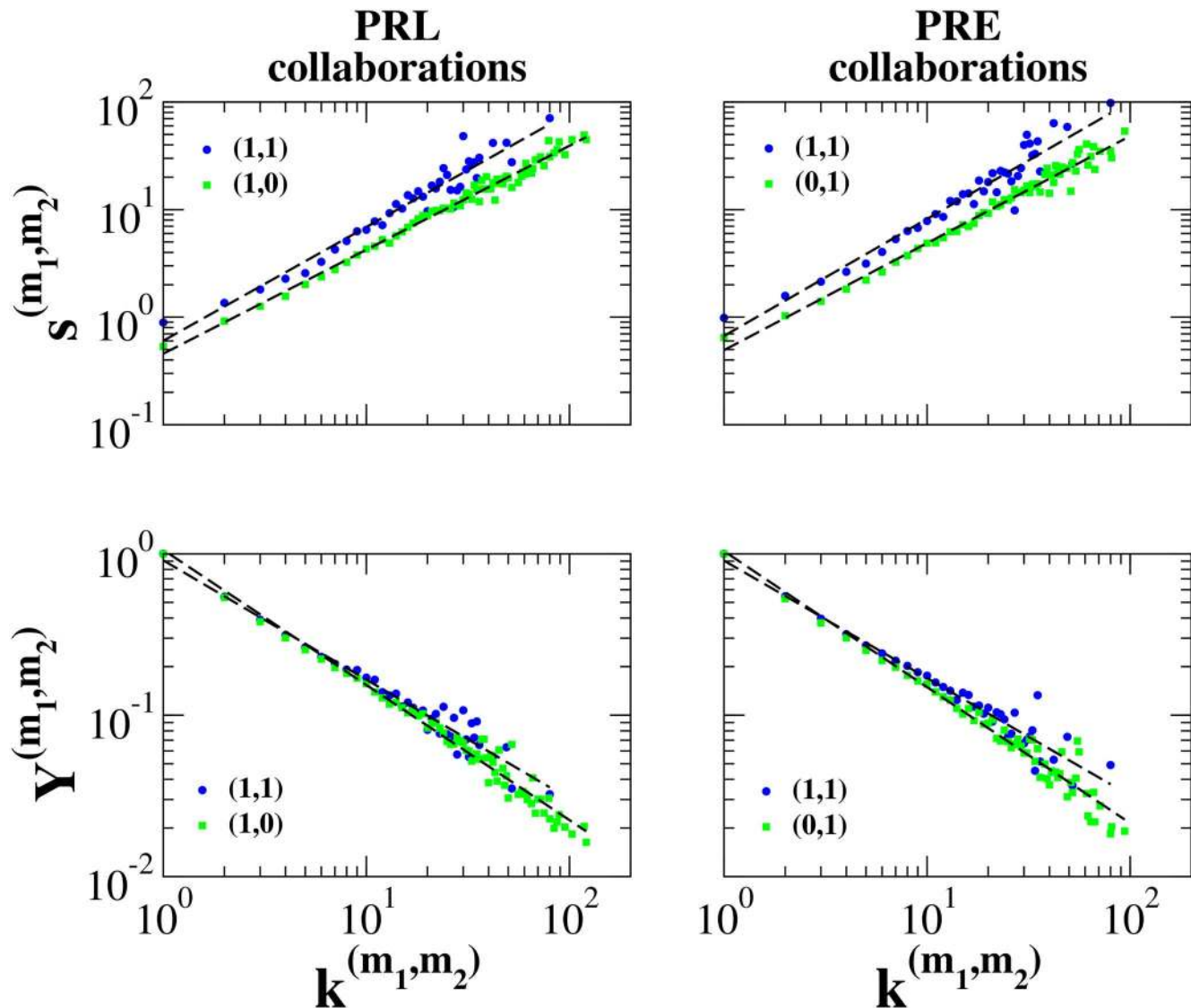
We indicate the degree of node  $i$  in layer  $\alpha$  with  $k_i^\alpha$ , defined as  $k_i^\alpha = \sum_{j=1}^N \theta(a_{ij}^\alpha)$ , where function  $\theta(x) = 1$  if  $x > 0$ , and  $\theta(x) = 0$  otherwise. In complex weighted networks, weights can be distributed across links more or less heterogeneously. A way to evaluate this heterogeneity is to introduce local properties such as the *strength*  $s_i^\alpha$  [21] and the *inverse participation ratio*  $Y_i^\alpha$  of node  $i$  in layer  $\alpha$  [22,23]:

$$s_i^\alpha = \sum_{j=1}^N a_{ij}^\alpha,$$

$$Y_i^\alpha = \sum_{j=1}^N \left( \frac{a_{ij}^\alpha}{s_i^\alpha} \right)^2. \quad (1)$$

As with single networks, in any given layer  $\alpha$ , the strength  $s_i^\alpha$  of a node indicates the sum of the weights of the links incident upon node  $i$  in layer  $\alpha$ , whereas the inverse participation ratio  $Y_i^\alpha$  indicates how unevenly the weights of the links of node  $i$  are distributed in layer  $\alpha$ . The inverse of  $Y_i^\alpha$  characterizes the effective number of links of node  $i$  in layer  $\alpha$ . Indeed,  $(Y_i^\alpha)^{-1}$  is greater than one and smaller than the degree of node  $i$  in layer  $\alpha$ , i.e.,  $(Y_i^\alpha)^{-1} \in (1, k_i^\alpha)$ . Moreover, if the weights of the links of node  $i$  are distributed uniformly, i.e.,  $w_{ij}^\alpha = s_i^\alpha / k_i^\alpha$ , we have  $(Y_i^\alpha)^{-1} = k_i^\alpha$ . Conversely, if the weight of one link is much larger than the other weights, i.e.,  $w_{ir}^\alpha \gg w_{ij}^\alpha$  for every  $j \neq r$ , then  $(Y_i^\alpha)^{-1} = 1$ .

In network theory, it is common practice to evaluate the conditional means of the strength and of the inverse participation ratio of the weights of links against the degree of nodes [21–23]. In a multiplex network, we will then consider the quantities  $s^\alpha(k) = \langle s_i^\alpha \delta(k_i^\alpha, k) \rangle$  and  $Y^\alpha(k) = \langle Y_i^\alpha \delta(k_i^\alpha, k) \rangle$ , where the average is calculated over all nodes with degree  $k$  in layer  $\alpha$ , and  $\delta(a, b)$  indicates the Kronecker delta. As in single networks [21],  $s_k^\alpha$  is expected to scale as



**Figure 2. Average multistrength and average inverse multiparticipation ratio versus multidegree in the CoCo-PRE/PRL multiplex network.** The average multistrengths and the average inverse multiparticipation ratios are fitted by a power-law distribution of the type described in Eq. (8) (fitted distributions are here indicated by black dashed lines). Statistical tests for the collaboration network of PRL suggest that the exponents  $\beta_{m,1}$  defined in Eq. (8) are the same, while exponents  $\lambda_{m,PRL}$  are significantly different. Similar results can be obtained for the exponents in the PRE collaboration layer. Nevertheless, multistrengths  $s^{(1,1),z}$  are always larger than multistrengths  $s^{(1,0),PRL}$  and  $s^{(0,1),PRE}$ , when multistrengths are calculated over the same number of multilinks, i.e.,  $k^{(1,1)} = k^{(1,0)} = k^{(0,1)}$  (see Text S1 for the statistical test on this hypothesis). doi:10.1371/journal.pone.0097857.g002

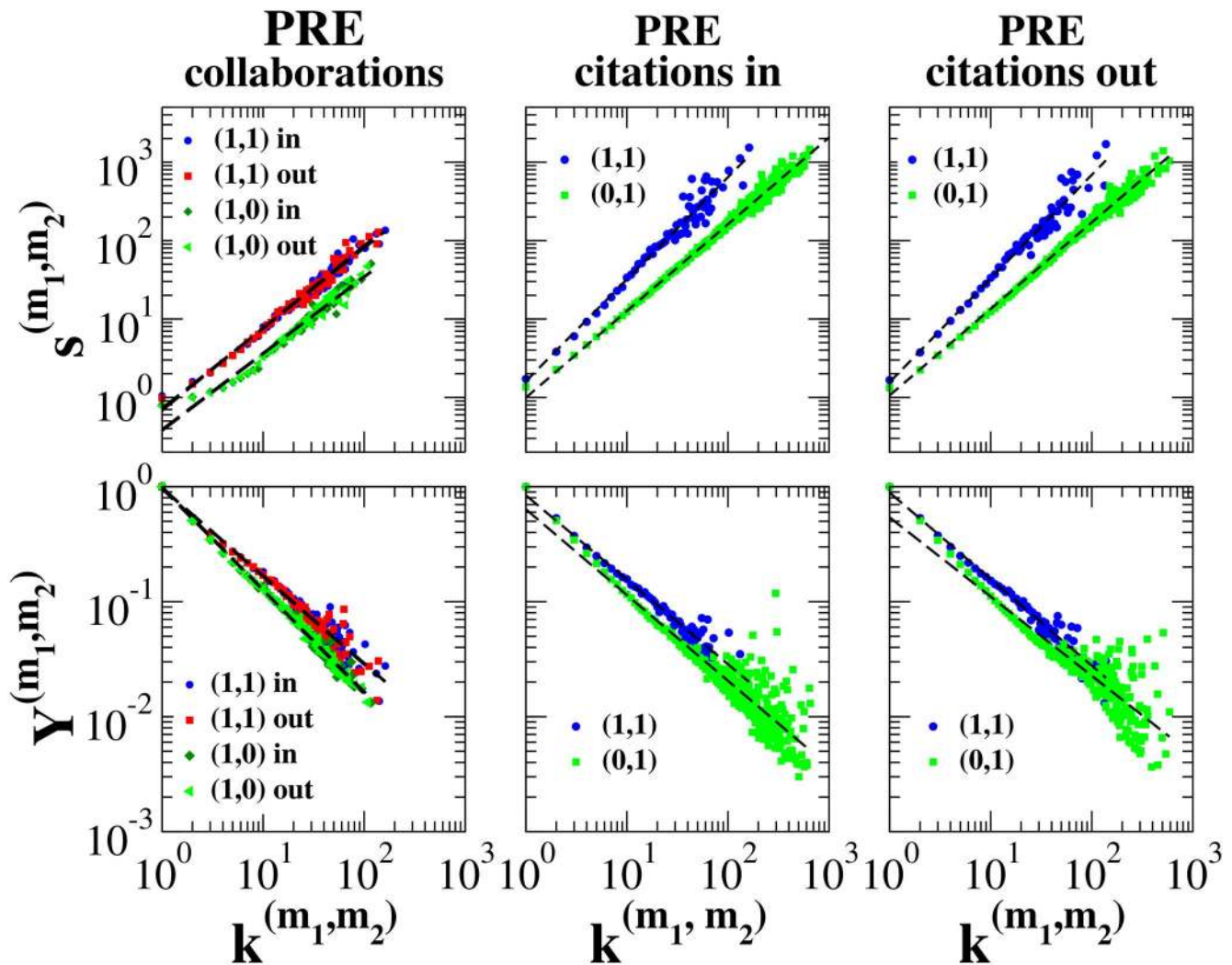
$$s^z(k) \propto k^{\beta_z}, \quad (2)$$

with  $\beta_z \geq 1$ . We can distinguish between two scenarios. In the first one, the average strength of nodes with degree  $k$  increases linearly with  $k$ , i.e.,  $\beta_z = 1$ . This indicates that, on average, the weights of the links incident upon the hubs do not differ from the weights of the links of less connected nodes. In the second scenario, the strength of the nodes with degree  $k$  increases super-linearly with  $k$ , i.e.,  $\beta_z > 1$ , thus indicating that, on average, the weights of the links incident upon the hubs are larger than the weights of the links of less connected nodes. In a multiplex network, it may be the case that weights are distributed in different ways across the layers. For instance, some layers may be characterized by a super-linear

growth of  $s_k^z$ , while other layers may show a linear dependence. Finally, the inverse participation ratio can be used in order to characterize the heterogeneity of the weights of the links incident upon nodes with a certain degree. In particular, it has been observed that, in many single weighted networks, the inverse participation ratio scales as an inverse power-law function of the degree of nodes. In a multiplex network, this would imply

$$Y^z(k) \propto \frac{1}{k^{\lambda_z}}, \quad (3)$$

where exponent  $\lambda_z \leq 1$  is layer-dependent.



**Figure 3. Properties of multilinks in the weighted CoCi-PRE multiplex network.** In the case of the collaboration network, the distributions of multistrengths versus multidegrees always have the same exponent, but the average weight of multilinks (1,1) is larger than the average weight of multilinks (1,0). Moreover, the exponents  $\lambda_{(1,0),col,in}$ ,  $\lambda_{(1,0),col,out}$  are larger than exponents  $\lambda_{(1,1),col,in}$ ,  $\lambda_{(1,1),col,out}$ . In the case of the citation layer, both the incoming multistrengths and the outgoing multistrengths have a functional behavior that varies depending on the type of multilink. Conversely, the average inverse multiparticipation ratio in the citation layer does not show any significant change of behavior when compared across different multilinks.

doi:10.1371/journal.pone.0097857.g003

### 2.3 Multilink, Multistrength, and Inverse Multiparticipation Ratio

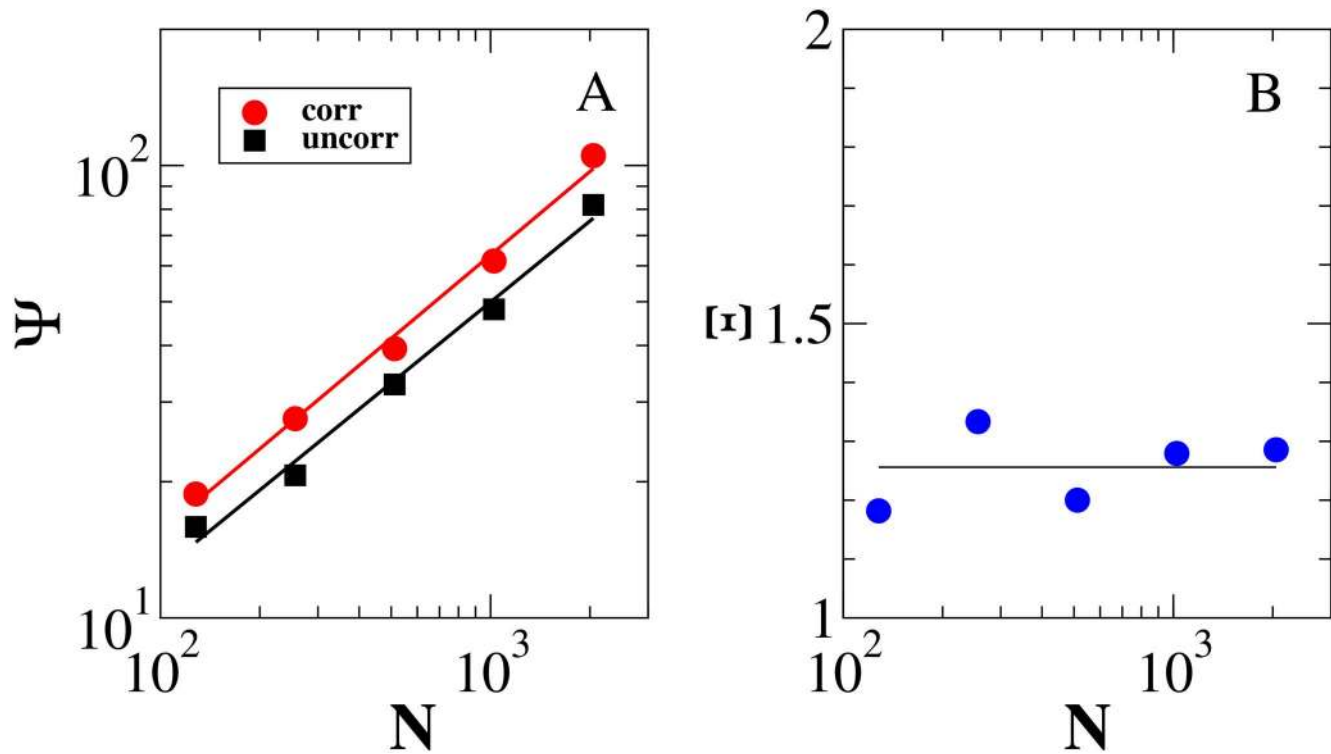
A number of multiplex networks are characterized by a significant overlap of links across the different layers [6,7]. In order to generalize the notion of overlap to weighted multiplex networks, in what follows we will draw on the concept of multilink [25]. Let us consider the vector  $\vec{m} = (m_1, m_2, \dots, m_x, \dots, m_M)$ , in which every element  $m_x$  can take only two values  $m_x = 0, 1$ . We define a *multilink*  $\vec{m}$  as the set of links connecting a given pair of nodes in the different layers of a multiplex network, and connecting them in the generic layer  $\alpha$  only if  $m_\alpha = 1$ . In particular, any two nodes  $i$  and  $j$  are always linked by a single multilink of type  $\vec{m} = \vec{m}^{ij} = (\theta(a_{ij}^1), \theta(a_{ij}^2), \dots, \theta(a_{ij}^M))$ , where  $\theta(x) = 1$  if  $x > 0$ , and  $\theta(x) = 0$  otherwise. The multilink  $\vec{m} = \vec{0}$  between two nodes represents the situation in which in all the layers of the multiplex network the two nodes are not directly linked.

We can now introduce the multiadjacency matrices  $\mathbf{A}^{\vec{m}}$  with elements  $A_{ij}^{\vec{m}}$  equal to 1 if there is a multilink  $\vec{m}$  between node  $i$  and node  $j$  and zero otherwise. In terms of the weighted adjacency matrices  $\mathbf{a}^\alpha$  of the multiplex network, the elements  $A_{ij}^{\vec{m}}$  of the multiadjacency matrix  $\mathbf{A}^{\vec{m}}$  are given by

$$A_{ij}^{\vec{m}} = \prod_{\alpha=1}^M [\theta(a_{ij}^\alpha) m_\alpha + (1 - \theta(a_{ij}^\alpha))(1 - m_\alpha)], \quad (4)$$

where  $\theta(x) = 1$  if  $x > 0$ , otherwise  $\theta(x) = 0$ . Even though there are  $2^M$  multiadjacency matrices, only  $2^M - 1$  of them are independent because the normalization condition,  $\sum_{\vec{m}} A_{ij}^{\vec{m}} = 1$ , must be satisfied for any pair of nodes  $i$  and  $j$ . Based on multi-adjacency matrices, we can define the *multidegree*  $k_i^{\vec{m}}$  of node  $i$  as





**Figure 4. (A) Value of the indicator  $\Psi$  defined in Eq. (10)** indicating the amount of information carried by the correlated and the uncorrelated multiplex ensembles of  $N$  nodes with respect to a null model in which the weights are distributed uniformly over the multiplex network. (B) Value of the indicator  $\Xi$  defined in Eq. (12) indicating the additional amount of information encoded in the properties of multilinks in the correlated multiplex ensemble with respect to the corresponding uncorrelated multiplex ensemble. The solid line refers to the average value of  $\Xi$  over the different multiplex network sizes.

doi:10.1371/journal.pone.0097857.g004

$$k_i^{\vec{m}} = \sum_{j=1}^N A_{ij}^{\vec{m}}, \quad (5)$$

which indicates how many multilinks  $\vec{m}$  are incident upon node  $i$ .

To study weighted multiplex networks, we now introduce two new measures. For layer  $\alpha$  associated to multilinks  $\vec{m}$ , such that  $m_\alpha > 0$ , we define the multistrength  $s_{i,\alpha}^{\vec{m}}$  and the inverse multiparticipation ratio  $Y_{i,\alpha}^{\vec{m}}$  of node  $i$ , respectively, as

$$s_{i,\alpha}^{\vec{m}} = \sum_{j=1}^N a_{ij}^\alpha A_{ij}^{\vec{m}}, \quad (6)$$

$$Y_{i,\alpha}^{\vec{m}} = \sum_{j=1}^N \left( \frac{a_{ij}^\alpha A_{ij}^{\vec{m}}}{\sum_r a_{ir}^\alpha A_{ir}^{\vec{m}}} \right)^2. \quad (7)$$

Since multistrength  $s_{i,\alpha}^{\vec{m}}$  can be non-zero only if  $m_\alpha = 1$ , for each layer  $\alpha$  the number of non-trivial multistrengths is  $2^{M-1}$ , and therefore the number of multistrengths that can be defined in a multiplex network of  $M$  layers is  $K = M2^{M-1}$ . Similarly, the number of inverse multiparticipation ratios  $Y_{i,\alpha}^{\vec{m}}$  is given by  $K = M2^{M-1}$ . The average multistrength of nodes with a given multidegree, i.e.,  $s^{\vec{m},\alpha}(k^{\vec{m}}) = \langle s_i^{\vec{m},\alpha} \delta(k_i^{\vec{m}}, k^{\vec{m}}) \rangle$ , and the average

inverse multiparticipation ratio of nodes with a given multidegree, i.e.,  $Y^{\vec{m},\alpha}(k^{\vec{m}}) = \langle Y_i^{\vec{m},\alpha} \delta(k_i^{\vec{m}}, k^{\vec{m}}) \rangle$ , are expected to scale as

$$s^{\vec{m},\alpha}(k^{\vec{m}}) \propto (k^{\vec{m}})^{\beta_{\vec{m},\alpha}}$$

$$Y^{\vec{m},\alpha}(k^{\vec{m}}) \propto \frac{1}{(k^{\vec{m}})^{\lambda_{\vec{m},\alpha}}}, \quad (8)$$

with exponents  $\beta_{\vec{m},\alpha} \geq 1$  and  $\lambda_{\vec{m},\alpha} \leq 1$ . The use of multilinks  $\vec{m}$  to describe multiplex properties is numerically feasible if the number of layers is smaller than the number of nodes, i.e.,  $M \ll \log(N)$ . If this condition is not satisfied, then the following quantities can be measured: the *overlap multiplicity*,  $v(\vec{m}) = \sum_\alpha m_\alpha$ , which indicates that multilink  $\vec{m}$  connects two nodes through  $v(\vec{m})$  links;  $s^z(v) = \langle s_{i,\alpha}^{\vec{m}} \rangle_{v(\vec{m})=v}$ ; and  $Y^z(v) = \langle Y_{i,\alpha}^{\vec{m}} \rangle_{v(\vec{m})=v}$ , where  $v = 1, 2, \dots, M$ .

### Empirical Evidence of Weighted Properties of Multilinks

In this section, we will draw on the measures introduced above and provide empirical evidence that, in weighted multiplex networks, weights can be correlated with the multiplex structure in a non-trivial way. To this end, we analyze the bibliographic dataset that includes all articles published in the APS journals (i.e., Physical Review Letters, Physical Review, and Reviews of Modern Physics) from 1893 to 2009. Of these articles, the dataset includes their citations as well as the authors. Here, we restrict our study only to articles published either in Physical Review Letters (PRL)

or in Physical Review E (PRE) and written by ten or fewer authors,  $n_p \leq 10$ . We constructed multiplex networks in which the nodes are the authors and links between them have a two-fold nature: scientific collaborations with weights defined as in [28] (see Text S1), and citations with weights indicating how many times author  $i$  cited author  $j$ .

In particular, we created the following two duplex networks (i.e., multiplex networks with  $M=2$ ):

- 1. CoCo-PRL/PRE:** *collaborations among PRL and PRE authors.* The nodes of this multiplex network are the authors with articles published both in PRL and PRE (i.e., 16,207 authors). These nodes are connected in layer 1 through weighted undirected links indicating the strength of their collaboration in PRL (i.e., co-authorship of PRL articles). The same nodes are connected in layer 2 through weighted undirected links indicating the strength of their collaboration in PRE (i.e., co-authorship of PRE articles).
- 2. CoCi-PRE:** *collaborations among PRE authors and citations to PRE articles.* The nodes of this multiplex network are the authors of articles published in PRE (i.e., 35,205 authors). These nodes are connected in layer 1 through weighted undirected links indicating the strength of their collaboration in PRE (i.e., co-authorship of PRE articles). The same nodes are connected in layer 2 through weighted directed links indicating how many times an author (with articles in PRE) cited another author's work, where citations are limited to those made to PRE articles.

Both these multiplex networks show a significant overlap of links and a significant correlation between degrees of nodes as captured by the Pearson correlation coefficient  $\rho$  (see Text S1). This finding supports the hypothesis that the two layers in each of the multiplex networks are correlated. That is, the existence of a link between two authors in one layer is correlated with the existence of a link between the same authors in the other layer. Moreover, the multidegrees of the multiplex networks are broadly distributed, and the hubs in the scientific collaboration network tend to be also the hubs in the citation network (see Text S1).

In the case of the CoCo-PRL/PRE network, multilinks  $\vec{m}=(1,0)$ ,  $\vec{m}=(0,1)$  and  $\vec{m}=(1,1)$  refer to collaborations only in PRL, only in PRE, and in both PRL and PRE, respectively. Moreover, to distinguish between the weights used when evaluating multistrength, we have  $\alpha=PRL$  or  $\alpha=PRE$ . Results indicate that multistrength and the inverse multiparticipation ratio behave according to Eq. (8) (see Fig. 2). The difference between exponents  $\beta_{\vec{m},PRL}$  for  $\vec{m}=(1,0)$  and  $\vec{m}=(1,1)$  is not statistically significant. Nevertheless, there is a statistically significant difference between the average weights of multilinks (1,0) and (1,1) in the PRL layer. As to the inverse multiparticipation ratio, there is a significant variation in the exponents,  $\lambda_{(1,0),PRL}=0.84 \pm 0.03$  and  $\lambda_{(1,1),PRL}=0.74 \pm 0.05$  (see Fig. 2, bottom left panel). This suggests that the weights of the collaborative links between co-authors of both PRL and PRE articles are distributed more heterogeneously than the weights of collaborative links between co-authors of articles published only in PRL (see Text S1 for details on the statistical tests). Similar results were found for multistrengths evaluated in the PRE layer (see Fig. 2, right panels).

These findings clearly indicate that the partial analysis of individual layers would fail to uncover the fact that the average weight of the link between authors that collaborated both on PRL and PRE articles is significantly larger than the average weight of the link between authors that collaborated only on articles published in one journal. Moreover, the difference in functional

behavior of the multipartition ratio across layers could not be captured if layers were analyzed separately.

In the case of the CoCi-PRE network, there are even more significant differences between the properties of the multilinks than in the previous network. In the CoCi-PRE network the functional behavior of multistrength also depends on the type of multilink. Figure 3 shows the average multistrength in the CoCi-PRE network. To distinguish between the weights used to measure multistrength, we have layer  $\alpha=col$ , which refers to the collaboration network constructed on PRE articles, and layer  $\alpha=cit$ , which refers to the citation network between PRE articles, where a distinction is also made between incoming (*in*) and outgoing (*out*) links. First, in the scientific collaboration network, exponents  $\beta_{\vec{m},col}$  are not statistically different, but the average weight of multilink (1,1) is larger than the average weight of multilinks (1,0),*in* and (1,0),*out*. Moreover, exponents  $\lambda_{(1,0),col,in}$  and  $\lambda_{(1,0),col,out}$  are larger than exponents  $\lambda_{(1,1),col,in}$ ,  $\lambda_{(1,1),col,out}$ , indicating that the weights of authors' collaborative links with other cited/citing authors are distributed more heterogeneously than the weights of authors' collaborative links with other authors with whom there are no links in the citation network. Second, in the citation network multistrengths follow a distinct functional behavior depending on the different type of multilink, and are characterized by different  $\beta_{\vec{m},cit,in/out}$  exponents. In fact the fitted values of these exponents are given by  $\beta_{(1,1)cit,in}=1.30 \pm 0.07$ ,  $\beta_{(1,1)cit,out}=1.32 \pm 0.08$ ,  $\beta_{(0,1)cit,in}=1.11 \pm 0.01$ , and  $\beta_{(0,1)cit,out}=1.10 \pm 0.02$ . This implies that, on average, highly cited authors are cited by their co-authors to a much greater extent than is the case with poorly cited authors. A similar, though much weaker effect was also found for the citations connecting authors that are not collaborators. Furthermore, in the citation layer the inverse multiparticipation ratio for multilink (1,1) is always larger than the inverse multiparticipation ratio for multilinks (1,0) and (0,1) (see Text S1 for details on the statistical test). Finally, when single layers were analyzed separately, we found  $\beta_{col}=1.03 \pm 0.04$  in the collaboration network, and  $\beta_{cit,in}=1.13 \pm 0.02$  and  $\beta_{cit,out}=1.14 \pm 0.03$  in the citation network. This indicates that in the citation network strength grows super-linearly as a function of degree, i.e., weights are not distributed uniformly. Nevertheless, correlations between weights and types of multilinks cannot be captured if the two individual layers are studied separately.

### 3.1 Assessing the Informational Content of Weighted Multilinks

Recent research on single networks has shown that the entropy of network ensembles provides a very powerful tool for quantifying their complexity [31–34]. Here, we propose a theoretical framework based on the entropy of multiplex ensembles for assessing the amount of information encoded in the weighted properties of multilinks. Multiplex weighted network ensembles can be defined as the set of all weighted multiplex networks satisfying a given set of constraints, such as the expected degree sequence and the expected strength sequence in every layer of the multiplex network, or the expected multidegree sequence and the expected multistrength sequence. A set of constraints imposed upon the multiplex network ensemble uniquely determines the probability  $P(\vec{G})$  of the multiplex networks in the ensemble (see Materials and Methods). The entropy  $\mathcal{S}$  of the multiplex ensemble can be defined in terms of  $P(\vec{G})$  as

$$\mathcal{S} = - \sum_{\vec{G}} P(\vec{G}) \log P(\vec{G}), \quad (9)$$

where  $\mathcal{S}$  indicates the logarithm of the typical number of multiplex networks in the ensemble. The smaller the entropy, the larger the amount of information stored in the constraints imposed on the network. The entropy can be regarded as an unbiased way to evaluate the informational value of these constraints.

In order to gauge the information encoded in a weighted multiplex network with respect to a null model, we define the indicator  $\Psi$ , which quantifies how much information is carried by the weight distributions of a weighted multiplex ensemble. In particular,  $\Psi$  compares the entropy of a weighted multiplex ensemble  $\mathcal{S}$  with the entropy of a weighted multiplex ensemble in which the weights are distributed homogeneously. Therefore,  $\Psi$  can be defined as

$$\Psi = \frac{|\mathcal{S} - \langle \mathcal{S} \rangle_{\pi(w)}|}{\langle (\delta \mathcal{S})^2 \rangle_{\pi(w)}}, \quad (10)$$

where  $\langle (\delta \mathcal{S})^2 \rangle_{\pi(w)}$  is the standard deviation, and the average  $\langle \dots \rangle_{\pi(w)}$  is calculated over multiplex networks with the same structural properties but with weights distributed homogeneously. In particular, when the weight distribution is randomized, the multiplex networks are constrained in such a way that each link must have a minimal weight (i.e.,  $w_{ij} \geq 1$ ), while the remaining of the total weight is distributed randomly over the links. In all the considered network ensembles we have assumed that the weights of the links can only take values that are multiple of a minimal weight. This assumption is by no means a limitation of this approach because for every finite network, there is always a minimal weight in the network such that this hypothesis is verified.

In order to evaluate the amount of information encoded in the weight of links in single layers and compare it to the information supplied by multistrength, we consider the following undirected multiplex ensembles:

- **Correlated weighted multiplex ensemble.** In this ensemble, we fix the expected multidegree sequence  $\{k_i^{\bar{m}}\}$ , and we set the expected multistrength sequence  $\{s_i^{\bar{m},\alpha}\}$  to be

$$s_i^{\bar{m},\alpha} = c_{\bar{m},\alpha} (k_i^{\bar{m},\alpha})^{\gamma_{\bar{m},\alpha}} \quad (11)$$

for every layer  $\alpha$ . We call  $\Psi^{corr}$  the  $\Psi$  calculated from this ensemble.

- **Uncorrelated weighted multiplex ensemble.** In this ensemble, we set the expected degree  $k_i^\alpha$  of every node  $i$  in every layer  $\alpha=1,2$  to be equal to the sum of the multidegrees (with  $m_\alpha=1$ ) in the correlated weighted multiplex ensemble. We set the expected strengths  $s_i^\alpha$  of every node  $i$  in every layer  $\alpha$  to be equal to the sum of the multistrengths of node  $i$  in layer  $\alpha$  in the correlated weighted multiplex ensemble. We call  $\Psi^{uncorr}$  the  $\Psi$  calculated from this ensemble.

In the correlated weighted multiplex ensemble the properties of the multilinks are accounted for, while in the uncorrelated weighted multiplex ensemble the different layers of the multiplex networks are analyzed separately (see Text S1 for the details). Finally, to quantify the additional amount of information carried by the correlated multiplex ensemble with respect to the uncorrelated multiplex ensemble, we define the indicator  $\Xi$  as

$$\Xi = \frac{\Psi^{corr}}{\Psi^{uncorr}}. \quad (12)$$

As an example of a possible application of the indicator  $\Xi$ , we focus on a case inspired by the CoCi-PRE multiplex network, where we consider different exponents  $\beta_{\bar{m},\alpha,in/out}$  for different multilinks. First, we created the correlated multiplex ensemble with power-law multidegree distributions  $P(k^{\bar{m}}) = C(k^{\bar{m}})^{-\gamma_{\bar{m}}}$  with exponents  $\gamma_{(1,m_2)} = 2.6$  for  $m_2=0,1$  and  $\gamma_{(0,1),(in/out)} = 1.9$  (where for multidegree  $(0,1)$  we imposed a structural cut-off). Multistrengths satisfy Eq. (11), with  $c_{\bar{m},\alpha} = 1$  and  $\beta_{(1,m_2),1} = 1$ , for  $m_2=0,1$ ;  $\beta_{(1,1),2} = 1.3$ , and  $\beta_{(0,1),2} = 1.1$ . Second, for the second layer, we created the uncorrelated version of the multiplex ensemble which is characterized by a super-linear dependence of the average strength on the degree of the nodes. We then measured  $\Psi$  as a function of network size  $N$  for these different ensembles. Numerically, the average  $\langle \dots \rangle_{\pi(w)}$  was evaluated from 100 randomizations. Figure 4 shows that  $\Psi$  increases with network size  $N$  as a power law, and that  $\Xi$  fluctuates around an average value of 1.256. These findings indicate that a significant amount of information is contained in multistrength and cannot be extracted from individual layers separately. Similar results, not shown here, were obtained with a correlated weighted multiplex ensemble characterized by non-trivial inverse multiparticipation ratios.

## Conclusions

In this paper, we have shown that weighted multiplex networks are characterized by significant correlations across layers, and in particular that weights are closely correlated with the multiplex network structure. To properly detect these correlations, we introduced and defined two novel weighted properties of multiplex networks, namely multistrength and the inverse multiparticipation ratio, that cannot be reduced to the properties of single layers. These weighted multiplex properties capture the crucial role played by multilinks in the distribution of weights, i.e., the extent to which there is a link connecting each pair of nodes in every layer of the multiplex network. To illustrate an example of weighted multiplex networks displaying non-trivial correlations between weights and topology, we analyzed the weighted properties of multilinks in two multiplex networks constructed by combining the co-authorship and citation networks involving the authors included in the APS dataset. Finally, based on the entropy of multiplex ensembles, we developed a theoretical framework for evaluating the information encoded in weighted multiplex networks, and proposed the indicator  $\Psi$  for quantifying the information that can be extracted from a given dataset with respect to a null model in which weights are randomly distributed across links. Moreover, we proposed a new indicator  $\Xi$  that can be used to evaluate the additional amount of information that the weighted properties of multilinks provide over the information contained in the properties of single layers. In summary, in this paper we have provided compelling evidence that the analysis of multiplex networks cannot be simplified to the partial analysis of single layers, and in particular that non-trivial information can be uncovered only by shifting emphasis on a number of weighted properties of multilinks.

## Materials and Methods

We can build a multiplex ensemble by maximizing the entropy  $\mathcal{S}$  of the ensemble given by Eq. (9) under the condition that the constraints imposed upon the multiplex networks are satisfied on average over the ensemble (soft constraints). We assume there are

$K$  of such constraints determined by the conditions.

$$\sum_{\vec{G}} P(\vec{G}) F_{\mu}(\vec{G}) = C_{\mu}, \quad (13)$$

for  $\mu = 1, 2, \dots, K$ , where  $F_{\mu}(\vec{G})$  determines one of the structural constraints that we want to impose on average on the multiplex network. The most unbiased multiplex ensemble satisfying the constraints given by Eqs. (13) maximizes the entropy  $\mathcal{S}$  under these constraints. In this ensemble, the probability  $P(\vec{G})$  for a multiplex network  $\vec{G}$  of the ensemble is given by

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\mu} \omega_{\mu} F_{\mu}(\vec{G}) \right], \quad (14)$$

where the normalization constant  $Z$  is called the “partition function” of the canonical multiplex ensemble, and is fixed by the normalization condition imposed on  $P(\vec{G})$ , whereas  $\omega_{\mu}$  are the Lagrangian multipliers enforcing the constraints in Eq. (13). The values of the Lagrangian multipliers  $\omega_{\mu}$  are determined by

## References

- Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74: 47–97.
- Newman MEJ (2003) Structure and function of complex networks. *SIAM Review* 45: 167–256.
- Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks: Structure and dynamics. *Physics Reports* 424: 175–308.
- Fortunato S (2010) Community detection in graphs. *Physics Reports* 486: 75–174.
- Kivelä M, Arenas A, Barthélemy M, Gleeson JP, Moreno Y, et al. (2013) Multilayer networks. arXiv: 1309.7233.
- Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. *PNAS* 107: 13636–13641.
- Cardillo A, Gómez-Gardeñes J, Zanin M, Romance M, Papo D, et al. (2013) Emergence of network features from multiplexity. *Sci. Rep.* 3: 1344.
- Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10: 186–198.
- Donges J, Schultz H, Marwan N, Zou Y, Kurths J (2011) Complex networks in climate dynamics. *Eur. Phys. Jour. B* 84: 635–651.
- Morris RG, Barthélemy M (2012) Transport on coupled spatial networks. *Phys. Rev. Lett.* 109: 128703 (2012)
- Battiston F, Nicosia V, Latora V (2014) Structural measures for multiplex networks. *Phys. Rev. E* 89: 032804.
- Halu A, Mondragón RJ, Panzarasa P, Bianconi G (2013) Multiplex PageRank. *PLoS ONE* 8: e78293.
- Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J-P (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328: 876–878.
- Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464: 1025–1028.
- Baxter GJ, Dorogovtsev SN, Goltsev AV, Mendes JFF (2012) Avalanche collapse of interdependent networks. *Phys. Rev. Lett.* 109: 248701.
- Gómez S, Díaz-Guilera A, Gómez-Gardeñes J, Pérez-Vicente CJ, Moreno Y, et al. (2013) Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.* 110: 028701.
- Brummitt CD, D’Souza RM, Leicht EA (2012) Suppressing cascades of load in interdependent networks. *PNAS* 109: E680–E689.
- Granovetter MS (1973) The strength of weak ties. *American Journal of Sociology* 78: 1360–1380.
- Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, et al. (2007) Structure and tie strengths in mobile communication networks. *PNAS* 104: 7332–7336.
- Karsai M, Perra N, Vespignani A (2013) The emergence and role of strong ties in time-varying communication networks. arXiv: 1303.5966.
- Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *PNAS*, 101: 3747–3752.
- Barthélemy M, Gondran B, Guichard E (2003). Spatial structure of the Internet traffic. *Physica A* 319: 633–642.
- Almaas E, Kovacs B, Vicsek T, Oltvai ZN, Barabási AL (2004). Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* 427: 839–843.
- Serrano MA, Boguñá M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. *PNAS* 106: 6483–6488.
- Bianconi G (2013) Statistical mechanics of multiplex networks: Entropy and overlap. *Phys. Rev. E* 87: 062806.
- Redner S (1998) How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* 4: 131–134.
- Newman MEJ (2001) The structure of scientific collaboration networks. *PNAS* 98: 404–409.
- Newman MEJ (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* 64: 016132.
- Radicchi F, Fortunato S, Markines B, Vespignani A (2009) Diffusion of scientific credits and the ranking of scientists. *Phys. Rev. E* 80: 056103.
- Radicchi F, Fortunato S (2008) Universality of citation distributions: Toward an objective measure of scientific impact. *PNAS* 105: 17266–172172.
- Park J, Newman MEJ (2004) Statistical mechanics of networks. *Phys. Rev. E* 70: 066117.
- Johnson S, Torres JJ, Marro J, Muñoz MA (2010) Entropic origin of disassortativity in complex networks. *Phys. Rev. Lett.* 104: 108702.
- Bianconi G, Pin P, Marsili M (2009) Assessing the relevance of node features for network structure. *PNAS* 106: 11433–11438.
- Sagarra O, Vicente CP, Díaz-Guilera A (2013) Statistical mechanics of multiedge networks. *Phys. Rev. E* 88: 062806.