

Multiplex PageRank

Arda Halu¹, Raúl J. Mondragón², Pietro Panzarasa³ Ginestra Bianconi^{4,*},

1 Department of Physics, Northeastern University, Boston, MA, USA

2 School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom

3 School of Business and Management, Queen Mary University of London, London, United Kingdom

4 School of Mathematical Sciences, Queen Mary University of London, London, United Kingdom

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

Abstract

Many complex systems can be described as multiplex networks in which the same nodes can interact with one another in different layers, thus forming a set of interacting and co-evolving networks. Examples of such multiplex systems are social networks where people are involved in different types of relationships and interact through various forms of communication media. The ranking of nodes in multiplex networks is one of the most pressing and challenging tasks that research on complex networks is currently facing. When pairs of nodes can be connected through multiple links and in multiple layers, the ranking of nodes should necessarily reflect the importance of nodes in one layer as well as their importance in other interdependent layers. In this paper, we draw on the idea of biased random walks to define the Multiplex PageRank centrality measure in which the effects of the interplay between networks on the centrality of nodes are directly taken into account. In particular, depending on the intensity of the interaction between layers, we define the Additive, Multiplicative, Combined, and Neutral versions of Multiplex PageRank, and show how each version reflects the extent to which the importance of a node in one layer affects the importance the node can gain in another layer. We discuss these measures and apply them to an online multiplex social network. Findings indicate that taking the multiplex nature of the network into account helps uncover the emergence of rankings of nodes that differ from the rankings obtained from one single layer. Results provide support in favor of the salience of multiplex centrality measures, like Multiplex PageRank, for assessing the prominence of nodes embedded in multiple interacting networks, and for shedding a new light on structural properties that would otherwise remain undetected if each of the interacting networks were analyzed in isolation.

Introduction

Despite recent advances [1–4] in the analysis of complex networks, a number of areas of investigation concerned with the description, prediction, and control of the dynamics of a variety of systems, including weather networks [5], social networks [6], and the brain [7], still remain largely unexplored. A large number of these systems cannot be properly understood unless they are regarded as components of higher-level systems in which various networks are connected with one another through a complex pattern of interdependencies [6, 8–10]. These higher-level systems can thus be seen as networks in which the interacting nodes are networks themselves that are characterized by their own structure and function, and co-evolve over time according to various patterns. Examples of such networks of networks include multimodal transportation networks, social networks, climatic systems, economic markets, energy-supply networks, and the human brain. For instance, the same individuals, groups and organizations can play different roles within a social relationship or can be linked through different types of social relationships (e.g., family relationships, acquaintanceship, friendship, and professional collaboration) [11–15], can have different affiliations [16, 17], and can communicate with one another using different technologies, such as

mobile phone, chat, e-mail, or video conferences [18, 19]. Each of these roles, relationships, affiliations, and communication technologies can in turn be associated with a different social network in which links between nodes refer to a distinct form of social interaction between the connected nodes. The same people, groups or organizations that interact in many different ways can thus be represented as the nodes of multiple co-evolving social networks that are themselves connected with one another as the various forms of social interaction affect one another over time [20–23]. Despite the ubiquity of these co-evolving and interconnected forms of interaction, so far network scientists have focused primarily on network datasets that contain only one type of social relation, thus neglecting the complexity of the connections between the various networks in which the same people interact. To uncover the nature and full breadth of social interaction, a special emphasis should be placed precisely on the structure and dynamics of the network of interacting social networks.

The system in which the same nodes belong to multiple interacting and co-evolving networks is typically referred to as a multiplex network or multigraph [24, 25]. In recent literature, there has been an upsurge of interest in multiplex networks. In particular, scholars have concentrated on the structural properties [6, 9, 10, 23, 26] and the antecedents [27, 28] of these networks, have shed light on diffusion processes [29, 30], cooperation [31, 32], exchange relations [33, 34], percolation phase transitions [8, 35], cascades [36], epidemic spreading [37], and election processes [38] occurring on them, and have developed modeling frameworks [39–41] and game-theoretic perspectives [32]. Among the structural properties of multiplex networks that scholars have only recently begun to address [42, 43], a crucial role is played by the centrality of nodes. In a multiplex network, the importance of a node depends on the connectivity patterns within and across the different layers of the network. For univariate networks in which no more than one link can connect the same pairs of nodes, a number of measures are available for assessing the importance of nodes. Over recent years these measures have become increasingly popular and salient for a variety of empirical domains. Among these measures, in this paper we concentrate our attention on PageRank, a centrality measure that has been successfully used not only for ranking web pages [44], but also for ranking scientists in citation networks [45] or species in food webs [46]. While PageRank was originally proposed as a centrality measure for univariate networks [44], its extension to multiplex networks remains largely unexplored. In particular, when the same pairs of nodes can be connected through multiple links co-evolving in multiple layers, a non-trivial problem is concerned with how to extend PageRank so as to capture the degree to which the ranking of nodes in one layer can affect, and be affected by, the ranking of the same nodes in other layers. This paper attempts to address this problem by proposing a generalization of PageRank to the case of multiplex networks.

To evaluate the relative popularity of a node in a network, PageRank centrality draws on the idea of a web surfer that visits different parts of the WWW at random. The random walker follows two strategies: the first is to jump to a node selected uniformly at random; the second is to jump, still randomly, to one of the walker’s neighbors. The popularity of a node is a function of the frequency with which the random walker visits the node. This frequency is then compared with the frequencies associated with all other nodes in the network. The ranking of nodes obtained according to these frequencies of being visited is precisely the ordering produced by the PageRank centrality measure, and reflects the relative popularity that each node has across the whole network.

To extend the PageRank centrality measure to the case of multiplex networks, we assume that the centrality a node has in one layer affects the centrality the node can obtain in another layer. This interplay between layers has a two-fold nature. First, the importance of a node in one layer may simply contribute to an increase in the node’s importance in another layer. Second, a node’s importance in one layer can amplify the node’s ability to derive benefits from the importance of other nodes that point to it in another layer. Alternatively, from the perspective of a biased random walk on complex networks, Multiplex PageRank can be described in terms of the bias that one layer exerts on the random jumps that a surfer makes in another layer [47, 48]. In this paper, we identify four versions of the Multiplex PageRank centrality measure, depending on how layers affect each other or, alternatively, exert a bias

upon the random jump. First, if the bias lies in the jump the random walker makes to any other node in the network, we obtain a measure that we call *Additive* Multiplex PageRank. Second, if the bias is exerted upon the jump the walker makes to any of its neighboring nodes, we obtain the *Multiplicative* Multiplex PageRank. The third variant is motivated by the fact that it is possible to have a bias in both jumps. In this case, we obtain the *Combined* Multiplex PageRank. Finally, the *Neutral* Multiplex PageRank refers to the case in which there is no bias in either jump, and thus the measure reduces to PageRank based on one single layer. For uncorrelated networks, we show how these centrality measures are correlated with the structural properties of the networks.

To clarify the meaning of the four versions of Multiplex PageRank, we apply these measures to a multiplex network formed by the juxtaposition of two networks, and show that the centrality of a node in one network depends on the centrality of the same node in the other network. Our application is concerned with online communication and is based on a multiplex network in which the same users can interact by sending instant messages to one another and by posting messages to a forum. As users can send messages directly to one another and at the same time participate in discussion groups within a forum, they can be regarded as embedded in two related online social networks. Our results show that the Multiplicative Multiplex PageRank of users displays a broad distribution, and is thus able to capture the emergence of high-ranked nodes, unlike what can be obtained through the application of the PageRank centrality measure to a single network.

Results

Introduced as a centrality measure for assessing the “importance” of web pages, the PageRank x_i of a node i in a network with N nodes is defined as [44]

$$x_i = \alpha_A \sum_j A_{ij} \frac{x_j}{g_j} + (1 - \alpha_A) \frac{1}{N}, \quad (1)$$

where A_{ij} are the elements of the adjacency matrix that are equal to one if node j points to node i and zero otherwise, $g_j = \max(1, k_j^{\text{out}}) = \max(1, \sum_r A_{rj})$, and $\alpha_A > 0$ is called the damping factor. PageRank can be interpreted as the stationary distribution of a random walk with additional random jumps. A random walker on site j jumps to one of j 's k_j^{out} out-neighbors with probability α_A , and to any other site chosen uniformly at random with probability $1 - \alpha_A$. The PageRank of a node is large to the extent that many other nodes point to it. The PageRank of a node is therefore expected to increase as a function of the node's in-degree, and indeed in [49,50] it was shown that, for uncorrelated networks, the PageRank of nodes can be approximated by their in-degree. If nodes in uncorrelated networks are grouped into classes depending on their extended degrees $\mathbf{k} = (k^{\text{in}}, k^{\text{out}})$, then the average PageRank for nodes of degree-class \mathbf{k} is

$$\bar{x}(\mathbf{k}) = \alpha_A \frac{k^{\text{in}}}{\langle k^{\text{in}} \rangle N} + (1 - \alpha_A) \frac{1}{N}, \quad (2)$$

where the symbol $\langle \dots \rangle$ indicates the average over the N nodes of the network.

PageRank was originally proposed for ranking web pages in response to text queries, and for this reason it was formalized as a centrality measure for directed networks [44]. It is, however, possible to extend the original definition to the case of undirected networks. For these networks, PageRank is

$$x_i = \alpha_A \sum_j A_{ij} \frac{x_j}{g_j} + (1 - \alpha_A) \frac{1}{N}, \quad (3)$$

where $g_j = \max(1, k_j)$ and k_j is the degree of node j . For undirected networks, the average PageRank $\bar{x}(k)$ of a node with degree k is given by

$$\bar{x}(k) = \alpha_A \frac{k}{\langle k \rangle N} + (1 - \alpha_A) \frac{1}{N}. \quad (4)$$

Compared to univariate networks, multiplex networks enable nodes to be connected with one another through more than one type of links, and as such offer a richer and more detailed backdrop against which the structural position of nodes can be assessed. The extension of PageRank to multiplex networks is therefore expected to shed light on novel ways for measuring the importance of nodes that capture their embeddedness in multiple interrelated relations. A ranking of nodes can thus be obtained that is likely to differ from the one originating simply from the position of nodes in one single network.

Here we offer a generalization and extension of the PageRank measure that can be applied to any multiplex network dataset. The assumption underlying our proposed measure is that the centrality of a node in one network can be affected by the centrality of the same node in another network. For the sake of simplicity, we consider the case in which the multiplex network is organized into two layers: network A and network B. Our analysis can easily be generalized to multiplex networks with more than two layers (See Materials and Methods). We indicate with A_{ij} the elements of the adjacency matrix of network A, and with B_{ij} the elements of the adjacency matrix of network B. For network A, we evaluate PageRank $\mathbf{x} = \{x_1, \dots, x_N\}$ using Eq. (1) with the parameter $\alpha_A > 0$. We then express the *Multiplex PageRank centrality* $\mathbf{X} = \{X_1, \dots, X_N\}$ of the nodes in network B with respect to PageRank \mathbf{x} .

Formally, we define the Multiplex PageRank centrality X_i of node i as

$$X_i = \alpha_B \sum_j x_i^\beta B_{ij} \frac{X_j}{G_j} + (1 - \alpha_B) \frac{x_i^\gamma}{N \langle x^\gamma \rangle}, \quad (5)$$

where $G_j = \sum_r B_{rj} x_r^\beta + \delta(0, \sum_r B_{rj} x_r^\beta)$, $\delta(a, b)$ is the Kronecker delta, $\alpha_B > 0$ is small enough to guarantee that the relation can be satisfied, and the exponents β and γ are both greater than or equal to zero. The first term in Eq. (5) refers to the contribution to node i 's centrality that derives from the centrality of the nodes pointing to i in network B. Like with the ordinary PageRank measure, this contribution is inversely proportional to the out-degree of node i 's in-neighbors. However, unlike the ordinary measure, Eq. (5) enables this contribution to be also affected by the centrality that both node i and its in-neighbors in network B have in network A. This interplay between the two networks has a two-fold effect on a node's centrality. First, the extent to which node i can derive some advantage from the centrality of its in-neighbors in network B becomes more significant as the centrality of i in network A becomes larger. The more prominent a node is in one layer, the more likely it is that the node can attract and gain benefit from other important nodes in another layer. Second, the contribution of each in-neighbor j to i 's centrality in network B is discounted by dividing j 's centrality by the sum of the centralities that j 's out-neighbors in network B have in network A. In other words, the benefits node i can derive from the centrality of any in-neighbor j in network B are diluted to the extent that j in network B points to many other nodes that are associated with high centrality in network A. An important node in one layer can attract important nodes in a different layer, but the benefit that can be gained in so doing are mitigated if there are many other nodes that have a similar capacity of attraction.

The second term in Eq. (5) reflects the contribution to node i 's centrality in network B that derives from i 's centrality in network A. By adding this second term, nodes that are not able to attract important neighbors in network B, can still derive some advantage simply by being central in network A. In the extreme case, a node with a zero in-degree in network B can still be associated with a non-zero value of centrality if the node has a non-zero centrality in network A. The assumption underlying this component of centrality is that the importance of a node in one layer is positively affected by the importance that the same node has in another layer, regardless of the node's capacity to attract other important nodes in the former layer.

Alternatively, Multiplex PageRank can also be regarded as the stationary distribution of a random walk with additional biased jumps. With probability α_B , a random walker on site j jumps to site i , one of j 's k_j^{out} out-neighbors selected with probability proportional to x_i^β , and with probability $1 - \alpha_B$ jumps to site i chosen with probability proportional to x_i^γ .

In what follows, we identify four important limiting cases of the Multiplex PageRank measure:

- *Additive Multiplex PageRank* ($\beta = 0, \gamma = 1$):

$$X_i = \alpha_B \sum_j B_{ij} \frac{X_j}{G_j} + (1 - \alpha_B) \frac{x_i}{N\langle x \rangle}, \quad (6)$$

where $G_j = \max(1, \sum_r B_{rj})$. This refers to the case in which the effect of network A on network B is exerted simply by “adding” some value to the centrality the nodes have in network B in proportion to the centrality they have in network A. Here the interplay between networks does not imply that the importance a node has in one network affects the node’s ability to derive benefits from important nodes in another network. Simply, being central in network A enables a node to gain more centrality in network B, regardless of the node’s capacity to attract important others in network B. Recast in terms of a random walk, this version of Multiplex PageRank refers to the case of a biased random walk, where the bias lies in the random jump to any node in network B. In particular, nodes with high PageRank in network A are preferred over other nodes with low PageRank in the same network as the destination of the random jumps that the walker makes in network B. A similar version of this PageRank measure, in which the random jumps are biased according to some predetermined distribution called “personalized vector”, has already been suggested in the computer science literature [51]. In qualitative agreement with this version, here we propose to regard a node’s PageRank in one layer as the node’s “personalized vector” in another layer.

- *Multiplicative Multiplex PageRank* ($\beta = 1, \gamma = 0$):

$$X_i = \alpha_B \sum_j x_i B_{ij} \frac{X_j}{G_j} + (1 - \alpha_B) \frac{1}{N}, \quad (7)$$

where $G_j = \sum_r B_{rj} x_r + \delta(0, \sum_r B_{rj} x_r)$. This refers to the case in which the effect of network A on network B lies in “multiplying” the benefits that a node gains from the importance of its in-neighbors in network B by a factor that is proportional to the node’s importance in network A. Thus, all benefits that can be obtained by being central in network A are contingent upon the connections that a node receives from important nodes in network B. The more important a node is in network A, the more value the node can extract from the connections received from important others in network B. Unlike the Additive version, the Multiplicative Multiplex PageRank does not enable a node to derive any added benefit in network B by simply being important in network A, regardless of the importance of the node’s in-neighbors in network B. Alternatively, this version of the measure also refers to the case of a biased random walk, where the bias lies in the walker’s choice of the out-neighbor as the destination of the jump. In particular, neighbors with high PageRank in network A are preferred over other neighbors with low PageRank in the same network.

- *Combined Multiplex PageRank* ($\beta = \gamma = 1$):

$$X_i = \alpha_B \sum_j x_i B_{ij} \frac{X_j}{G_j} + (1 - \alpha_B) \frac{x_i}{N\langle x \rangle}, \quad (8)$$

where $G_j = \sum_r B_{rj} x_r + \delta(0, \sum_r B_{rj} x_r)$. This refers to the case in which the effect of network A on network B lies in “combining” the additive and multiplicative benefits a node in network B can gain by being central in network A. In this case, a node’s high centrality in network A can boost its centrality in network B both in itself and at the same time by amplifying the node’s ability to derive centrality from other important nodes. Alternatively, this version of Multiplex PageRank refers to the case in which both the destination of the random jump and the selection of the random walker’s out-neighbor in network B are biased in that they favor nodes with high PageRank in network A over nodes with low PageRank in the same network.

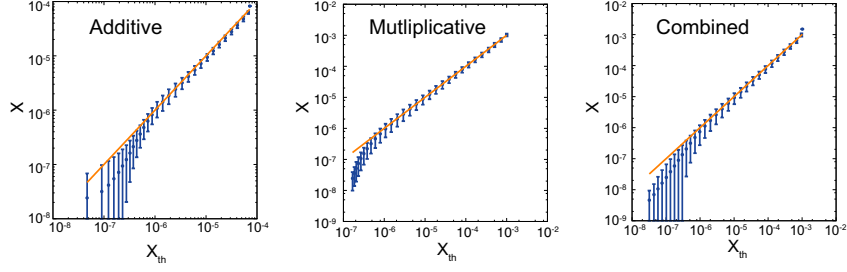


Figure 1. Data X versus theory X_{th} for the Additive, Multiplicative, and Combined versions of PageRank. In each of the three panels, the PageRank of the data is plotted against the corresponding value obtained through our theoretical approximation. Multiplex PageRank was evaluated using an iterative procedure with the standard values $\alpha_A = \alpha_B = 0.85$. The accuracy of the algorithm was set at 10^{-11} .

- *Neutral Multiplex PageRank* ($\beta = \gamma = 0$):

$$X_i = \alpha_B \sum_j B_{ij} \frac{X_j}{G_j} + (1 - \alpha_B) \frac{1}{N}, \quad (9)$$

where $G_j = \max(1, \sum_r B_{rj})$. This refers to the case in which there is no effect of network A upon network B, and thus Multiplex PageRank reduces to the PageRank based simply on network B in isolation.

Clearly these limiting cases can be generalized so as to be applied also to a multiplex network that combines two undirected networks or a directed network and an undirected one. Moreover, the above definitions can be further generalized so as to accommodate cases in which the rankings \mathbf{x} are obtained using different centrality measures, such as the eigenvector centrality.

Following [49,50], we performed a mean-field calculation of the average Multiplex PageRank $\overline{X}(\mathbf{k}_B, x)$ of a node with degree $\mathbf{k}_B = (k_B^{\text{in}}, k_B^{\text{out}})$ in network B and PageRank x in network A. We define $\overline{X}(\mathbf{k}_B, x)$ in the following way

$$\overline{X}(\mathbf{k}_B, x) = \frac{1}{NP(\mathbf{k}_B, x)} \sum_{i|\mathbf{k}_B, i=x} X_i, \quad (10)$$

where $P(\mathbf{k}_B, x)$ is the probability that a node has degree $\mathbf{k}_B = (k_B^{\text{in}}, k_B^{\text{out}})$ in network B and PageRank x in network A. In particular, performing a mean-field calculation (see Materials and Methods) valid for an uncorrelated network B, we obtain

$$\overline{X}(\mathbf{k}_B, x) = \alpha_B x^\beta k_B^{\text{in}} \frac{1}{\langle x^\beta k_B^{\text{in}} \rangle N} + (1 - \alpha_B) \frac{x^\gamma}{N \langle x^\gamma \rangle}. \quad (11)$$

To verify the validity of Multiplex PageRank, we generated a duplex network with 10^7 nodes and 8×10^7 links in each layer. In both layers, the in- and out-degrees decay as a power law k^η , where $\eta_A^{\text{out}} = 2.8$, $\eta_A^{\text{in}} = 2.1$ and $\eta_B^{\text{out}} = 2.5$, $\eta_B^{\text{in}} = 2.5$. The dependence of the Additive, Multiplicative, and Combined versions of PageRank upon in-degree is shown in Fig. 1. For small values of PageRank, the deviation from the diagonal is due to large fluctuations of PageRank in correspondence of small values of in-degree, as was also observed by Fortunato et al. [50].

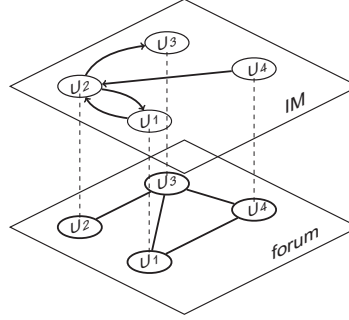


Figure 2. Sketch of the multiplex online social network in which users communicate by exchanging instant messages and by posting messages to a forum.

Discussion

We apply the Multiplex PageRank measure to the multiplex network created from an online community at the University of California, Irvine [52]. The multiplex network includes two layers. The first layer corresponds to a directed instant messaging (IM) network in which a directed link is established from one user to another if the former sends one or more online instant messages to the latter. The second layer is a bipartite network in which a link is established between a user and a discussion group of a forum when the former posts a message to the latter. While the IM network dataset covers the period from April 19 to October 26, 2004, the forum became active at a later time when users were already communicating through instant messages. The bipartite network thus covers a more restricted period than the IM network, from May 14 to October 26, 2004. The two networks also differ in the number of users: the total number of active users recorded for the IM network is 1,899, of whom only 899 posted at least one message in the forum. Moreover, users that were active in the forum created 552 thematic groups, each aimed at the discussion of a specific topic.

The analysis of the multiplex network covers the restricted observation period beginning on June 4, 2004, when both networks were operational and exhibited a fairly stable pattern of activity. At any specific day, and with a daily frequency, we constructed the instantaneous cumulative networks reflecting all the social interactions that took place in the three weeks' period ending on that day. Measurements thus create a time series with 124 sample networks starting on June 25, 2004. The multiplex network can be represented by the juxtaposition of the two time-varying adjacency matrices $\mathbf{A}(t)$ and $\mathbf{B}(t)$ that describe the IM network and the one-mode projection of the bipartite forum network, respectively (see Fig. 2). In particular, the adjacency matrix $\mathbf{A}(t)$ describes directed links between users, i.e. $A_{ij}(t) = 1$ if user j sent at least one message to user i in a given time window. For the forum network, the adjacency matrix $\mathbf{B}(t)$ describes an undirected and unweighted network between the users of the forum, where $B_{ij}(t) = 1$ if both user i and user j posted at least one message to a common discussion group in a given time window.

The application of Multiplex PageRank to online communication is motivated by the fact that users can enhance their ranking by engaging in multiple and interrelated ways of communication. In our specific case, users' prominence in the IM network (\mathbf{A}) is likely to have an impact upon the prominence they gain by communicating and interacting in the forum network (\mathbf{B}). To fully capture this intertwined nature of users' prominence, we begin by calculating each user's PageRank x_i based on the IM network. The Multiplex PageRank X_i of user i in the forum is then obtained by expressing the PageRank user i has in the forum as a function of the user's PageRank x_i based on the IM network. Formally, the Multiplex

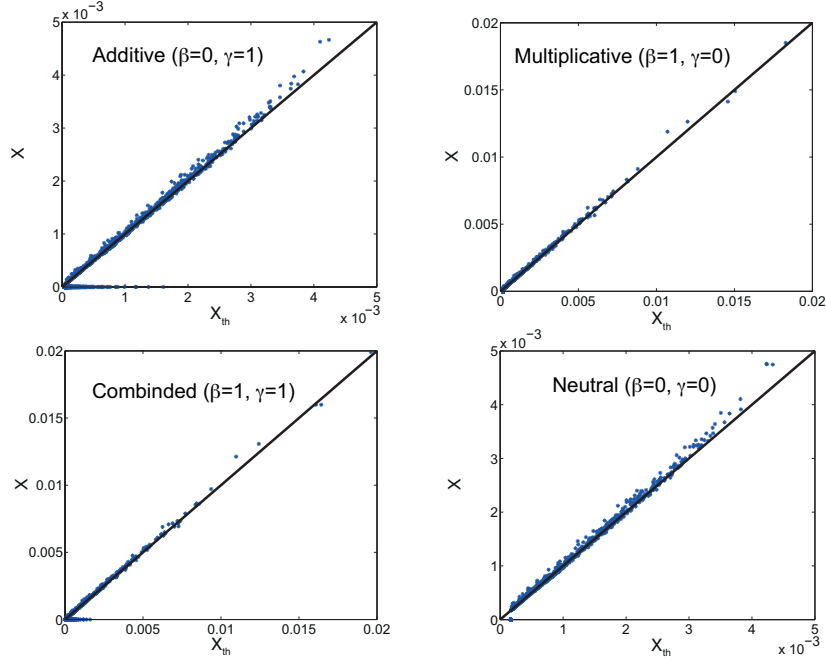


Figure 3. The Additive, Multiplicative, Combined, and Neutral users' Multiplex PageRanks X plotted against the mean field expectation X_{th} (solid line) for the IM-forum multiplex network dataset. The damping factors used for the IM and forum data are $\alpha_A = \alpha_B = 0.85$.

PageRank of user i in the forum at time t is

$$X_i(t) = \alpha_B \sum_j [x_i(t)]^\beta B_{ij}(t) \frac{X_j(t)}{G_j(t)} + (1 - \alpha_B) \frac{[x_i(t)]^\gamma}{\langle [x(t)]^\gamma \rangle_B N_F(t)}, \quad (12)$$

where $N_F(t)$ is the number of active users in the forum at time t , $\langle \dots \rangle_B$ denotes the average of $x(t)$ based only on the nodes that belong to network B at time t , and $G_j(t) = \sum_r B_{rj}(t) [x_r(t)]^\beta + \delta(0, \sum_r B_{rj}(t) [x_r(t)]^\beta)$. In the above formula, the PageRank x_i of node i in the IM network at time t is given by

$$x_i(t) = \alpha_B \sum_j A_{ij}(t) \frac{x_j(t)}{g_j(t)} + (1 - \alpha_B) \frac{1}{N}, \quad (13)$$

with $g_j(t) = \max(1, k_j^{A,out}(t))$. For the IM-forum multiplex network, we compared the values of Multiplex PageRank with the theoretical expectations obtained in the case of an uncorrelated network. We found a very good agreement between the two sets of values (see Fig. 3).

A crucial issue affecting a large number of applications, ranging from the online sale of books to usage of Twitter tags, that rely on measures for ranking items is concerned with the stability of the rankings over time [53]. To address this problem, here we investigate the stability of the top-ranked users in the forum, and compare the rankings obtained using the different proposed versions of Multiplex PageRank.

In order to evaluate the stability of rankings in our dataset, we select the top five users with the highest Multiplex PageRank at the end of the whole observation period ($t = 124$), and track their evolution over

time. In Fig. 4 we plot the time evolution of the values of PageRank X of the five users with the highest Additive, Multiplicative, Combined, and Neutral PageRanks. Users are ranked in decreasing order, from top to bottom. The figure indicates that the top five users with the highest values of Multiplicative Multiplex PageRank and Combined Multiplex PageRank are the same, and the top five users with the highest values of Additive and Neutral Multiplex PageRank are the same (with the exception of user 297 and user 511; note that user 297 has extremely high PageRank x in the IM network). The Neutral Multiplex PageRank refers to the case in which $\beta = 0$ and $\gamma = 0$ in Eq. (5), and thus produces a ranking of users that coincides with the one obtained by taking into consideration only users' position in the forum.

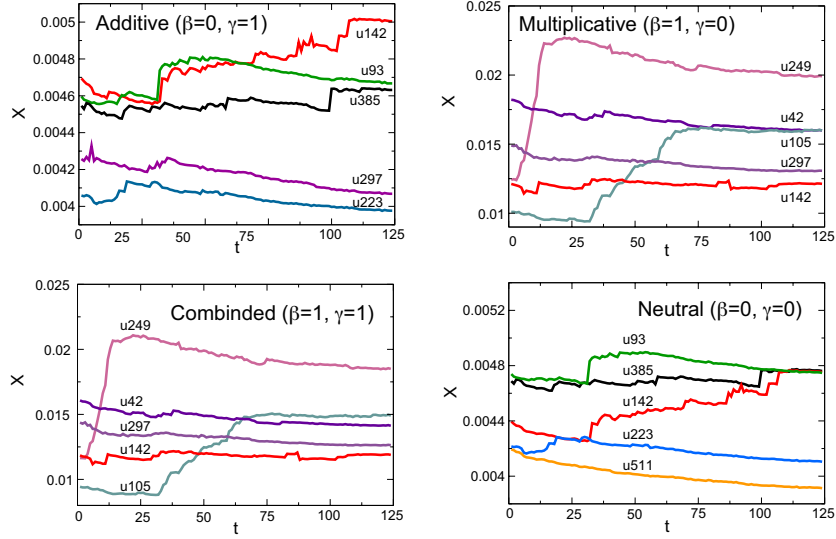


Figure 4. The time evolution of the values of Additive, Multiplicative, Combined, and Neutral Multiplex PageRank X for the 5 top-ranked users. The damping factors used for the IM and forum network data are $\alpha_A = \alpha_B = 0.85$. Each time step reflects the cumulative interactions in a three-week time window.

When the importance of the users in the forum is adjusted to also reflect their position in the IM network simply by adding a bias in the random jump to any node, as occurs with the Additive Multiplex PageRank, the identity of the top five users in the resulting ranking does not change significantly. On the contrary, the introduction of a bias in the random walker's choice of the out-neighbor as the destination of the jump, as occurs with the Multiplicative Multiplex PageRank, is responsible for a substantial change in the ranking of users. In particular, the emergence of new top-ranked users when the Multiplicative Multiplex PageRank is adopted suggests that there are synergies between the activities of these users in the IM and forum networks. The way they communicate and rise to prominence in one network affects how they communicate and rise to prominence in the other network. Thus, taking into account the multiplex nature of the dataset helps unveil these synergies and the multi-faceted nature of users' prominence that would otherwise remain undetected if only one layer were investigated. Moreover, Fig. 4 shows that the Combined Multiplex PageRank, by adding a bias both in the random jump to any node and in the walker's choice of the out-neighbor, does not produce any substantial change in the ordering of the top-ranked users with respect to the ranking that is obtained with the Multiplicative Multiplex PageRank.

In most applications, the use of PageRank for assessing the importance of nodes is aimed primarily at

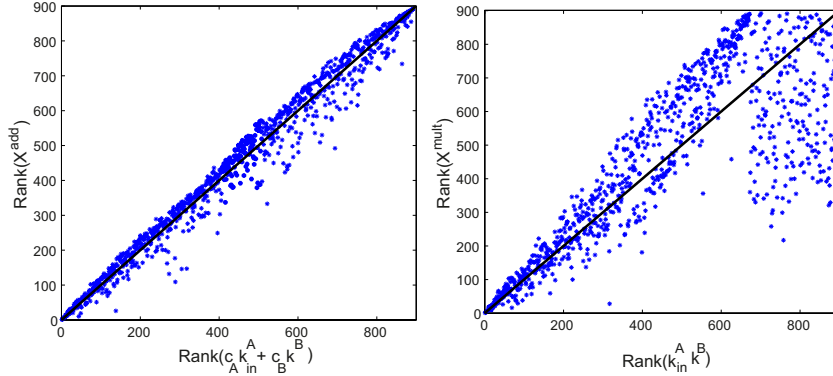


Figure 5. The estimated ranks of users according to the sum $(c_A k_{in}^A + c_B k^B)$ and product $(k_{in}^A k^B)$ of their in-degrees and degrees plotted against the values of their Additive and Multiplicative Multiplex PageRank, respectively. In the figure $c_A = (1 - \alpha_B)\alpha_A / (\langle k_{in}^A \rangle N \langle x \rangle_B)$ and $c_B = \alpha_B / \langle k^B \rangle$. Note that the node with rank 1 is the most important node of the network, and therefore the Additive and Multiplicative Multiplex PageRanks of the most important nodes of the online social network are correlated, respectively, with a linear combination or the product of the users' in-degrees and degrees.

producing a ranking of nodes rather than associate each of them with a specific value of centrality. As in a variety of networks nodes' PageRank is closely related to their in-degree, especially for nodes with high in-degree, it has become common practice to use in-degree as a proxy for PageRank. In the case of the multiplex online social network, drawing on our theoretical framework, we tested the hypothesis that the Additive Multiplex PageRank correlates with a linear combination of nodes' in-degree in network A and their degree in network B, while the Multiplicative PageRank correlates with the product between nodes' in-degree in network A and their degree in network B. Fig. 5 does indeed provide support in favor of our hypothesis. Findings thus suggest that the Additive and Multiplicative versions of PageRank can be well approximated by the following two simple measures of centrality for multiplex networks: respectively, the linear combination of nodes' degrees in the different layers, and the multiplication of nodes' degrees in the different layers.

Finally, we found that in our dataset the distribution of Multiplex PageRank is broad, especially the one of the Multiplicative and Combined versions of Multiplex PageRank (See Figure ??), as is expected in the case of multiplex networks with positive correlations between degrees of nodes in the different layers.

In conclusion, in this paper we introduced Multiplex PageRank, namely a centrality measure that can be used to identify and rank important nodes in multiplex networks. In particular, we defined four versions of this measure: the Additive, Multiplicative, Combined, and Neutral Multiplex PageRank. We then analyzed how these measures correlate with the degree of the nodes in the different layers, both at the mean-field level and using data on an online social network. The empirical application of these measures to our dataset indicated that taking into consideration the multiplex nature of social interaction helps uncover the emergence of rankings of nodes and of structural properties that would otherwise remain undetected if only univariate single networks were investigated.

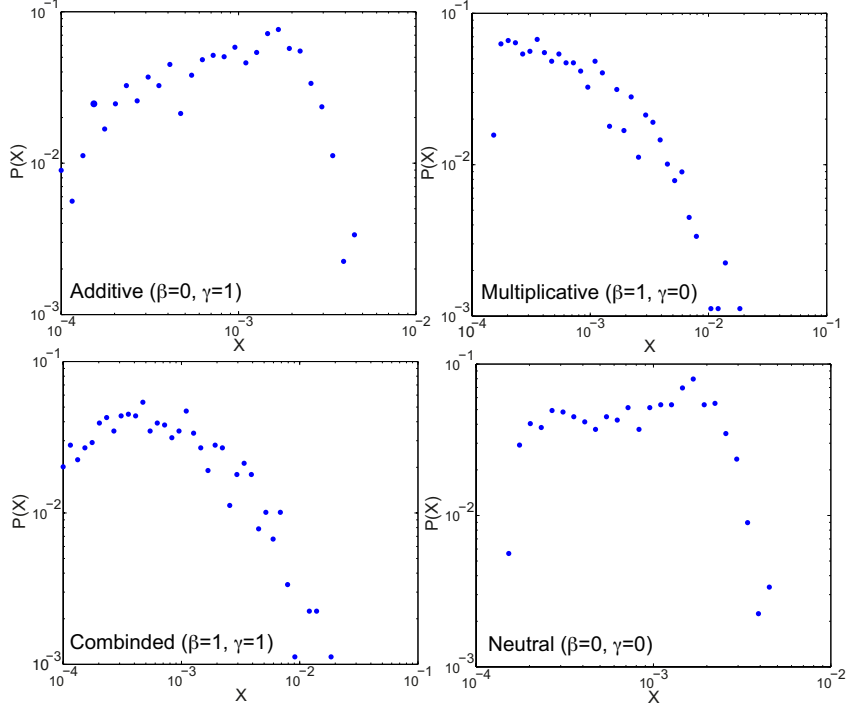


Figure 6. The distribution $P(X)$ of the Additive, Multiplicative, Combined, and Neutral versions of Multiplex PageRank X for users in the IM-forum multiplex network dataset. The damping factors used for the IM and forum data are $\alpha_A = \alpha_B = 0.85$.

Materials and Methods

The dataset used in this paper consists of communication records of anonymized users that were randomly assigned an identification number. This ensures privacy protection and compliance with ethical guidelines. The dataset can be obtained upon request from Pietro Panzarasa. The use of the dataset is free, provided the appropriate credit is given to the authors and a reference is made to this paper and to paper [52].

Approximating Multiplex PageRank by the degree. Derivation of Eq. (11).

In order to calculate the Multiplex PageRank X_i of node i , we use an iterative procedure. PageRank $\mathbf{X}^n = \{X_1^n, \dots, X_N^n\}$ at time step n can then be calculated from PageRank \mathbf{X}^{n-1} at time step $n-1$ according to the recursive equation

$$X_i^n = \alpha_B \sum_j x_i^\beta B_{ij} \frac{X_j^{n-1}}{G_j} + (1 - \alpha_B) \frac{x_i^\gamma}{\langle x^\gamma \rangle N}, \quad (14)$$

where PageRank X_i is given by $X_i = \lim_{n \rightarrow \infty} X_i^n$ and where $G_j = \sum_r B_{rj} x_r^\beta + \delta(0, \sum_r B_{rj} x_r^\beta)$.

Following [49, 50], we divide the PageRank of nodes in network B into different classes, where two or more nodes belong to the same class if they have the same in- and out-degree and the same PageRank x . We define $\bar{X}(\mathbf{k}^{(B)}, x)$ as the average value of the PageRank of nodes in the degree class $\mathbf{k}^{(B)} = (k_{\text{in}}^{(B)}, k_{\text{out}}^{(B)})$

in network B and with PageRank x in network A

$$\overline{X}(\mathbf{k}^{(B)}, x) = \frac{1}{NP(\mathbf{k}^{(B)}, x)} \sum_{i|\mathbf{k}_i^{(B)}=\mathbf{k}^{(B)}, x_i=x} X_i, \quad (15)$$

where $P(\mathbf{k}^{(B)}, x)$ is the probability that a node has degree $\mathbf{k}^{(B)}$ in network B and PageRank x in network A. Similarly, we define $\overline{X}^n(\mathbf{k}^{(B)}, x)$ from the iterative procedure

$$\begin{aligned} \overline{X}^n(\mathbf{k}^{(B)}, x) &= \frac{1}{NP(\mathbf{k}^{(B)}, x)} \sum_{i|\mathbf{k}_i^{(B)}=\mathbf{k}^{(B)}, x_i=x} X_i^n \\ &= \frac{\alpha_B}{NP(\mathbf{k}^{(B)}, x)} \left(\sum_{i|\mathbf{k}_i^{(B)}=\mathbf{k}^{(B)}, x_i=x} \sum_j x_i^\beta B_{ij} \frac{X_j^{n-1}}{G_j} \right) + (1 - \alpha_B) \frac{x^\gamma}{\langle x^\gamma \rangle N}. \end{aligned} \quad (16)$$

The term G_j in the above equation can be approximated as

$$G_j = \sum_r x_r^\beta B_{rj} + \delta(0, k_{j,\text{out}}^{(B)}) \simeq k_{\text{out},j}^{(B)} \sum_{\mathbf{k}'^{(B)}, x'} P_{\text{out}}^{(B)}(\mathbf{k}'^{(B)}, x' | \mathbf{k}_j^{(B)}, x_j) (x')^\beta + \delta(0, k_{j,\text{out}}^{(B)}), \quad (17)$$

where $P_{\text{out}}^{(B)}(\mathbf{k}'^{(B)}, x' | \mathbf{k}_j^{(B)}, x_j)$ is the probability of reaching a node with degree $\mathbf{k}'^{(B)}$ and PageRank x' by following a link in network B from a node of degree $\mathbf{k}_j^{(B)}$ and PageRank x_j . If the nodes belonging to class $\{\mathbf{k}_j^{(B)}, x_j\}$ are uncorrelated to the nodes of class $\{\mathbf{k}^{(B)}, x\}$, then

$$P_{\text{out}}^{(B)}(\mathbf{k}'^{(B)}, x' | \mathbf{k}_j^{(B)}, x_j) = \frac{k_{\text{in}}'^{(B)}}{\langle k_{\text{in}}^{(B)} \rangle} P(\mathbf{k}'^{(B)}, x'), \quad (18)$$

such that

$$G_j \simeq k_{\text{out},j}^{(B)} \frac{\langle x^\beta k_{\text{in}}^{(B)} \rangle}{\langle k_{\text{in}}^{(B)} \rangle} + \delta(0, k_{j,\text{out}}^{(B)}). \quad (19)$$

Using the approximation for G_j we can express the sum in Eq. (16) via a mean-field approximation, obtaining

$$\begin{aligned} \sum_{i|\mathbf{k}_i^{(B)}=\mathbf{k}^{(B)}, x_i=x} \sum_j x_i^\beta B_{ij} \frac{X_j^{n-1}}{G_j} &= \sum_{\mathbf{k}''} \sum_{x''} \sum_{i|\mathbf{k}_i^{(B)}=\mathbf{k}^{(B)}, x_i=x} \sum_{j|\mathbf{k}_j^{(B)}=\mathbf{k}'', x_j=x''} x_i^\beta B_{ij} \frac{X_j^{n-1}}{k_{\text{out},j}^{(B)}} \frac{\langle k_{\text{in}}^{(B)} \rangle}{\langle x^\beta k_{\text{in}}^{(B)} \rangle} \\ &\simeq \sum_{\mathbf{k}''} \sum_{x''} \sum_{i|\mathbf{k}_i^{(B)}=\mathbf{k}^{(B)}, x_i=x} \overline{X}^{n-1}(\mathbf{k}''^{(B)}, x'') \frac{\langle k_{\text{in}}^{(B)} \rangle}{\langle x^\beta k_{\text{in}}^{(B)} \rangle} \frac{x_i^\beta}{(k''^{\text{out}})^{(B)}} \sum_{j|\mathbf{k}_j^{(B)}=\mathbf{k}'', x_j=x''} B_{ij} \\ &\simeq \sum_{\mathbf{k}''} \sum_{x''} \sum_{i|\mathbf{k}_i^{(B)}=\mathbf{k}^{(B)}, x_i=x} \overline{X}^{n-1}(\mathbf{k}''^{(B)}, x'') \frac{\langle k_{\text{in}}^{(B)} \rangle}{\langle x^\beta k_{\text{in}}^{(B)} \rangle} \frac{x_i^\beta (k^{(B)})_{\text{in}}}{(k''^{\text{out}})^{(B)}} P_{\text{in}}^{(B)}(\mathbf{k}''^{(B)}, x'' | \mathbf{k}^{(B)}, x), \end{aligned}$$

where we used the mean-field approximation

$$\sum_{j|\mathbf{k}_j^{(B)}=\mathbf{k}'', x_j=x''} X_j^{n-1} B_{ij} \simeq \overline{X}^{n-1}(\mathbf{k}''^{(B)}, x'') \sum_{j|\mathbf{k}_j^{(B)}=\mathbf{k}'', x_j=x''} B_{ij} \quad (20)$$

and $P_{\text{in}}^{(B)}(\mathbf{k}''^{(B)}, x'' | \mathbf{k}^{(B)}, x)$ is the probability that, by following an incoming link of a node with degree $\mathbf{k}^{(B)}$ and PageRank x in network B, a predecessor of the node with degree $\mathbf{k}''^{(B)}$ and PageRank x'' can be reached. In an uncorrelated network, this quantity is given by

$$P_{\text{in}}^{(B)}(\mathbf{k}''^{(B)}, x'' | \mathbf{k}^{(B)}, x) = \frac{k_{\text{out}}''^{(B)}}{\langle k_{\text{in}}^{(B)} \rangle} P^{(B)}(\mathbf{k}''^{(B)}, x''). \quad (21)$$

Inserting Eqs. (20) and (21) in Eq. (16), and taking the limit $n \rightarrow \infty$, we obtain

$$\bar{X}(\mathbf{k}^{(B)}, x) = \alpha_B \frac{x^\beta k_{\text{in}}^{(B)}}{\langle x^\beta k_{\text{in}}^{(B)} \rangle N} + (1 - \alpha_B) \frac{x^\gamma}{N \langle x^\gamma \rangle}, \quad (22)$$

where the values of x can be approximated from Eq. (2)

$$\bar{x}(\mathbf{k}^{(A)}) = \alpha_A \frac{k_{\text{in}}^{(A)}}{\langle k_{\text{in}}^{(A)} \rangle N} + (1 - \alpha_A) \frac{1}{N}. \quad (23)$$

Extension of Multiplex PageRank to multiplex networks with more than two layers.

The proposed Multiplex PageRank centrality measure, presented in the main text for the specific case of a duplex network, can easily be extended to multiplex networks with more than two layers. Let us consider a multiplex network with M layers given in a predetermined order, where each layer $\ell = 1, 2, \dots, M$ corresponds to a network with adjacency matrix $A_{ij}^{(\ell)}$. We can define the Multiplex PageRank $X_i^{(\ell)}$ recursively in the following way. At the first level of the iteration $\ell = 1$, we have the single-layer PageRank $X_i^{(1)}$ defined as

$$X_i^{(1)} = \alpha^{(1)} \sum_j A_{ij}^{(1)} \frac{X_j^{(1)}}{G_j^{(1)}} + (1 - \alpha^{(1)}) \frac{1}{N}, \quad (24)$$

where $G_j^{(1)} = \max(1, \sum_r A_{rj})$. We then include the information about the structure of the other layers, and obtain

$$X_i^{(\ell)} = \alpha^{(\ell)} \sum_j \left[X_i^{(\ell-1)} \right]^\beta A_{ij}^{(\ell)} \frac{X_j^{(\ell)}}{G_j^{(\ell)}} + (1 - \alpha^{(\ell)}) \frac{\left[X_i^{(\ell-1)} \right]^\gamma}{N \langle [X^{(\ell-1)}]^\gamma \rangle}, \quad (25)$$

where $G_j^{(\ell)} = \sum_r A_{rj}^{(\ell)} \left[X_r^{(\ell-1)} \right]^\beta + \delta \left(0, \sum_r A_{rj}^{(\ell)} \left[X_r^{(\ell-1)} \right]^\beta \right)$. For the sake of simplicity, here we have chosen exponents β and γ that do not depend on ℓ , but in general it is also possible to consider the case in which the exponents β and γ are dependent on the layers ℓ .

Acknowledgments

References

References

1. Albert R, Barabási A L (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74: 47-97.
2. Barrat A, Barthélemy M, Vespignani A (2008) *Dynamical Processes on Complex Networks*. Cambridge: Cambridge University Press.
3. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) *Complex Networks: Structure and Dynamics*. *Physics Reports* 424:175-308.
4. Newman MEJ (2003) The structure and function of complex networks. *SIAM Review* 45: 167-256.
5. Donges J, Zou Y, Marwan N, Kurths J (2009) The backbone of the climate network. *EPL*, 87: 48007 (2009).
6. Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. *PNAS* 107:13636-13641.
7. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10: 186-198.
8. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464: 1025-1028.
9. Cardillo A, Gómez-Gardeñes J, Zanin M, Romance M, Papo D, et al. (2013) Emergence of network features from multiplexity. *Sci. Rep.* 3: 1344.
10. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328: 876-878.
11. Kapferer B (1969) Norms and the manipulation of relationships in a work context. In: Michell JC, editor. *Social Networks in Urban Situations*. Manchester: Manchester University Press.
12. Lomi A, Pattison PE (2006) Manufacturing relations: An empirical study of the organization of production across multiple networks. *Organization Science* 17(3): 313-332.
13. Uzzi B (1996) The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review* 61, 674-698.
14. Vebrugge LM (1979) Multiplexity in adult friendships. *Social Forces* 57(4): 1286-1309.
15. Zelizer VA (2005) *The Purchase of Intimacy*. Princeton, NJ: Princeton University Press
16. Lee S, Monge P (2011) The coevolution of multiplex communication networks in organizational communities. *Journal of Communication* 61: 758-779.
17. Wheeldon PD (1969) The operation of voluntary associations and personal networks in the political process of an inter-ethnic community. In: Mitchell JD, editor. *Social Networks in Urban Situations: Analyses of Personal Relationship in Central African Towns*. Manchester: Manchester University Press. pp. 128-180.

18. Baym NK, Zhang YB, Lin MC (2004) Social interactions across media. *New Media Soc* 6: 299-318.
19. Huang Y, Shen C, Williams D, Contractor N (2009) Virtually there: Exploring proximity and homophily in a virtual world. *Proc 2009 Int Conf Comput Sci Eng* 4: 354-359.
20. Gibbons, DE (2004). Friendship and advice networks in the context of changing professional values. *Administrative Science Quarterly*, 49(2): 238-262.
21. Ingram P, Zou X (2008) Business friendships. *Research in Organizational Behavior* 28: 167-84.
22. Lincoln JR, Miller J (1979). Work and friendship ties in organizations. Comparative analysis of relational networks. *Administrative Science Quarterly* 24(2): 181-199.
23. Maggioni MA, Breschi S, Panzarasa P (2013) Multiplexity, growth mechanisms and structural variety in scientific collaboration networks. *Industry and Innovation* 20(3): 185-194.
24. Flament C (1963) *Applications of Graph Theory to Group Structures*. Englewood Cliffs, NJ: Prentice-Hall.
25. Wasserman S, Faust K (1994) *Social Network Analysis. Methods and Applications*. Cambridge: Cambridge University Press.
26. Rank ON, Robins GL, Pattison PE (2010) Structural logic of intraorganizational networks. *Organization Science* 21(3): 745-764.
27. Ferriani S, Fonti F, Corrado R (2012) The social and economic bases of network multiplexity: Exploring the emergence of multiplex ties. *Strategic Organization* 11(1): 7-34.
28. Shipilov AV, Li SX (2012) The Missing link: The effect of customers on the formation of relationships among producers in the multiplex triads. *Organization Science* 23(2): 472-91.
29. De Domenico M, Sole A, Gómez S, Arenas A (2013) Random walk on multiplex networks. [arXiv:1306.0519](https://arxiv.org/abs/1306.0519).
30. 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.
31. Gimeno J, Woo CY (1996) Economic multiplexity: The structural embeddedness of cooperation in multiple relations of interdependence. In: Baum JAC, Dutton JE, Editors. *The Embeddedness of Strategy*. Greenwich, CT: JAI Press. pp. 323-361.
32. Gómez-Gardeñes J, Reinares I, Arenas A, Floria LM (2012) Evolution of cooperation in multiplex networks. *Sci. Rep.* 2: 620.
33. Kuwabara K (2011) Cohesion, cooperation, and the value of doing things together: How economic exchange creates relational bonds. *American Sociological Review* 76(4): 560-80.
34. Kuwabara K, Luo J, Sheldon O (2010) Multiplex exchange relations. In: Thye SR, Lawler EJ, Editors. *Advances in Group Processes* 27. Bingley, UK: Emerald Group Publishing Limited. pp. 239-268.
35. Zhao K, Bianconi G (2013) Percolation on interacting, antagonistic networks. *J. Stat. Mech.* P05005.
36. Brummitt CD, D'Souza RM, Leicht EA (2012) Suppressing cascades of load in interdependent networks. *PNAS* 109: E680-E689.

37. Saumell-Mendiola A, Serrano MA, Boguñá M (2012) Epidemic spreading on interconnected networks. *Phys. Rev. E* 86: 026106.
38. Halu A, Zhao K, Baronchelli A, Bianconi G (2013) Connect and win: The role of social networks in political elections. *EPL* 102: 16002.
39. Bianconi G (2013) Statistical mechanics of multiplex ensembles: Entropy and overlap. *Phys. Rev. E* 87: 062806.
40. Kim JY, Goh KI (2013) Coevolution and correlated multiplexity in multiplex networks. *Phys. Rev. Lett.* 111: 058702.
41. Nicosia V, Bianconi G, Latora V, Barthelemy M (2013) Growing multiplex networks. *Phys. Rev. Lett.* 111: 058701.
42. Markose S, Shaghghi AR (2013) Systemic importance in multi-layer financial networks: The Case of Derivative Markets. (private communication).
43. Solá L, Romance M, Criado R, Flores J, Garcia del Amo A, Boccaletti S (2013) Eigenvector centrality of nodes in multiplex networks. [arXiv:1305.7445](https://arxiv.org/abs/1305.7445)
44. Brin S, Page L (1998) The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems* 30: 107-117.
45. Liu X, Bollen J, Nelson L, Van de Sompel H (2005) Co-authorship networks in the digital library research community. *Information processing & management* 41: 1462-1480.
46. Allesina S, Mercedes P (2009) Googling food webs: can an eigenvector measure species' importance for coextinctions? *PloS Comp. Bio.* 5: e1000494.
47. Gómez-Gardeñes J, Latora V (2008) Entropy rate of diffusion processes on complex networks. *Phys. Rev. E* 78: 065102(R).
48. Siudem G, Holyst JA (2013) Diffusion and entropy production for multi-networks with fitness factors. [arXiv:1303.2650](https://arxiv.org/abs/1303.2650).
49. Fortunato S, Boguñá M, Flammini A (2008) Approximating PageRank from in-degree. *Lectures Notes in Computer Science* 4936: 59.
50. Fortunato S, Flammini A (2007) Random walks on directed networks: the case of PageRank. *Int. Jour. Bifurcation and Chaos* 17: 2343-2353.
51. Jeh G, Widom J (2003) Scaling personalized web search. *WWW '03 Proceedings of the 12th International Conference on World Wide Web*: 271-279.
52. Panzarasa P, Opsahl T, Carley KM (2009) Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology* 60(5): 911-932.
53. Blumm N, Ghoshal G, Forró Z, Schich M, Bianconi G, et al. (2012) Dynamics of ranking processes in complex systems. *Phys. Rev. Lett.* 109: 128701.