Link prediction in multiplex networks via triadic closure

Alberto Aleta ⬤,[1] Marta Tuninetti,[1] Daniela Paolotti,[1] Yamir Moreno,[2,3,1] and Michele Starnini[1,*]

[1]*ISI Foundation, via Chisola 5, 10126 Torino, Italy*
[2]*Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain*
[3]*Department of Theoretical Physics, University of Zaragoza, 50018 Zaragoza, Spain*

Link prediction algorithms can help to understand the structure and dynamics of complex systems, to reconstruct networks from incomplete data sets, and to forecast future interactions in evolving networks. Available algorithms based on similarity between nodes are bounded by the limited amount of links present in these networks. In this Rapid Communication, we reduce this latter intrinsic limitation and show that different kinds of relational data can be exploited to improve the prediction of new links. To this aim, we propose a link prediction algorithm by generalizing the Adamic-Adar method to multiplex networks composed by an arbitrary number of layers, that encode diverse forms of interactions. We show that this metric outperforms the classical single-layered Adamic-Adar score and other state-of-the-art methods, across several social, biological, and technological systems. As a by-product, the coefficients that maximize the multiplex Adamic-Adar metric indicate how the information structured in a multiplex network can be optimized for the link prediction task, revealing which layers are redundant. Interestingly, this effect can be asymmetric with respect to predictions in different layers. Our work paves the way for a deeper understanding of the role of different relational data in predicting new interactions and provides another algorithm for link prediction in multiplex networks that can be applied to a plethora of systems.

Network science has been established as a pivotal tool to characterize the structure of real-world complex systems, that involves multiple types of relations among their fundamental components [1]. One of the most important challenges within the complex systems framework is to elucidate which entities are related to which others and what are the types of these relationships [2]. Within the network science domain, this scientific task translates into a link prediction problem, that attempts to estimate the likelihood of the existence of a link between two nodes, based on the observed links and attributes of nodes [3,4]. Link prediction algorithms are extremely helpful in at least two directions: to reconstruct networks from incomplete data sets and to forecast future interactions in evolving networks. Examples of the first application can be found in biological networks, such as protein interaction networks, where many links are still unknown and their existence must be demonstrated by expensive experiments [5]. Prediction algorithms help in focusing experimental efforts toward those links most likely to exist. The second task, link forecasting, is routinely applied in online social networks, such as Facebook. New friendships are indeed recommended based on

link prediction algorithms, so that individuals can efficiently find peers they are interested in [6,7].

The link prediction problem is a long-standing challenge at the intersection between computer science and statistical physics communities. Traditional algorithms include Markov chains and statistical models [8], while recent approaches from the physics community, such as random walk processes and maximum-likelihood methods, have been considered [9,10]. Link prediction algorithms can be classified mainly into two categories: similarity-based methods and probabilistic models [11,12]. Since the latter can be computationally unfeasible for large networks, a lot of attention has been devoted to the creation of good similarity scores. Many of these similarity methods are based on the same basic idea, that two nodes are likely to be linked if they share a common neighbor [13,14]. Despite its simplicity, this concept has proven to be quite useful for highly assortative networks, such as scientific collaboration networks [15]. However, as signaled by Jia *et al.* [16], the prediction power of any similarity-based link prediction algorithm is bounded due to the limited amount of links present in the network.

In network science, richly structured data can be represented by multilayered networks, in which each layer accounts for a different type of interaction [17,18]. For instance, social interactions can have different purposes (e.g., leisure versus work) and happen through various communication channels, including face-to-face interactions, e-mail, Facebook, phone calls, and so on. The idea of predicting links in multilayer networks has been explored during the last decade from several different points of view. For instance, Davis *et al.* [19]

---

proposed a technique to include multirelational data for link prediction from a probabilistic point of view. Similarly, several extensions of probabilistic models to multilayer networks have been proposed [20–25]. Other works, rather than focusing on incorporating new data to already existing networks, used multilayer structures to focus on the temporal evolution of the networks [26,27]. Several studies extended the notion of neighborhood to multilayer networks [28–31], focusing on networks of two layers. However, a fundamental question is still unanswered: How can different kinds of relational data be exploited to improve the prediction of new interactions? For instance, to which extent are face-to-face interactions predictive of new Facebook friendships? Interestingly, it has been recently shown that the multiplex network representation can be redundant in some cases, as the information encoded in some layers can be effectively included in others, reducing the number of layers [32]. Therefore, how can link prediction algorithms optimize the information structured in a multiplex network representation, that can be suboptimally organized?

In this Rapid Communication, we address these questions by proposing a metric for link prediction in multiplex networks, based on a generalization of the Adamic-Adar method for single-layered networks [14]. Our metric fully exploits the complexity of the relationships that might be established across the fundamental components of complex systems, by considering all possible triadic closures in the corresponding multiplex representation. We show that this score, that can be applied to any multiplex topology composed by an arbitrary number of layers, is able to outperform other metrics based on single-layered similarity between nodes, across several social and biological systems. We show that the information encoded in different layers can be asymmetric with respect to the link prediction problem: For example, face-to-face interactions can be partially predictive of new Facebook friendships, but not vice versa.

We consider eight different data sets spanning several types of social, biological, and technological systems, represented as multiplex networks: (i) Copenhagen networks study (CNS), where four layers represent physical proximity, phone calls, text messages, and Facebook friendships among university students [33]. (ii) *C. elegans* genetic (CEG): Genetic and protein interactions of the *C. elegans*, where three layers represent direct, physical, and additive genetic interactions [34]. (iii) *C. elegans* neural (CEN): A neural network of the *C. elegans*, where three layers represent electric, chemical monoadic, and chemical polyadic interactions [35], (iv) CS-Aarhus (CSA): A social network of employees of the Computer Science Department at Aarhus, where five layers represent Facebook, leisure, work, coauthorship, and lunch interactions [36]. (v) CKM physicians (CKM): A social network of physicians, where three layers represent who they ask for advice, who they discuss cases with, and who are their friends [37]. (vi) EU air (EUA): An air transportation network of Europe, where 27 layers represent airlines routes [38]. (vii) Lazega (LAZ): A social network of partners and associates of a corporate law partnership, where three layers represent cowork, friendship, and advice [39]. (viii) Vickers (VIC): A social network of students in a school in Victoria, Australia, where three layers represent who they get on with, who are their

best friends, and who they prefer to work with [40]. See Table S1 of the Supplemental Material (SM) [41] for details about the data sets.

In the following, we will contrast different algorithms for link prediction on these data sets. The quality of link prediction algorithms can be evaluated by two metrics: the receiver operating characteristics (ROC) curve, with the corresponding area under the curve (AUC) value, and the precision. The precision can be computed as n^*/n, where n is the number of new links that we want to predict and n^* is the amount of correct predictions among the top n links. Thus, it provides complementary information to the one given by the AUC. It is important to highlight that, due to the limited amount of links present in a network, the AUC of any similarity-based link prediction algorithm is bounded [16]. For instance, if similarity is based on common neighbors, two nodes without any neighbor in common will have a score equal to zero. The number of scoreless links bounds the maximum and minimum values of the AUC to $\text{AUC}_{\min} = \frac{1}{2}(1 + p_1)(1 - p_2)$ and $\text{AUC}_{\max} = \text{AUC}_{\min} + p_1 p_2$, where p_1 (p_2) is the fraction of links with a score different from 0 among those links that will (will not) exist in the future (see Sec. 2 of the SM [41] for details). Note that only when $p_1 = p_2 = 1$, i.e., there are no scoreless links, it holds $\text{AUC}_{\min} = 0$ and $\text{AUC}_{\max} = 1$.

We propose a generalization of the Adamic-Adar (AA) score [14], one of the most common and successful methods for link prediction in social networks. The AA score between nodes u and v is given by the number of common neighbors weighted by their degree,

$$\text{AA}(u, v) = \sum_{w \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\ln(k_w)}, \qquad (1)$$

where $\Gamma(u)$ represents the set of neighbors of node u and $k_w = |\Gamma(w)|$ is the degree of node w. In a multiplex network, the AA score can be applied to different layers, depending on which layer α the set of neighbors $w \in \Gamma_\alpha(u) \cap \Gamma_\alpha(v)$ is considered, where $\Gamma_\alpha(u)$ represents the set of neighbors of node u in layer α.

For example, let us consider that we are interested in predicting future phone calls among the participants in the CNS. The classic AA method considers the set of neighbors in the same phone calls layer. If the AA score is applied to the layer representing Facebook friendship, instead, the rationale is that two individuals are more likely to interact offline (phone call) if they share many friends on Facebook (i.e., common neighbors in the Facebook layer). The same reasoning applies to other layers. Table I shows the precision and AUC values (together with its theoretical bounds) to predict phone calls, obtained for the AA method applied to each layer of the CNS (excluding physical proximity interactions for being much denser than others). Interestingly, while the maximum precision (0.04) is obtained by applying the AA score to the same calls layer, the maximum AUC (0.69) is obtained by considering the Facebook layer. This implies that this kind of interactions (Facebook friendship) include useful information to predict new links not encoded in the phone calls layer. This is also reflected in a larger maximum theoretical bound of the AUC for the Facebook layer with respect to the phone calls layer. Note also that by using the aggregated network, in

TABLE I. Precision and AUC to predict new phone calls in the CNS data set, obtained by using the classical Adamic-Adar metric on each layer (calls, Facebook, texts), and on the aggregated network. Theoretical bounds of the AUC are showed. Predictions are tested over the set of nonoverlapping links over all layers (7% of the total). Best results among the layers are highlighted in bold.

Method	Precision	AUC	AUC [worst-best]
Random	0	0.50	[0-1]
AA$_{calls}$	**0.04**	0.60	[0.60-0.60]
AA$_{facebook}$	0	**0.69**	[0.59-0.70]
AA$_{sms}$	0	0.60	[0.60-0.60]
AA$_{aggregated}$	0	0.76	[0.65-0.80]

which all layers are projected onto a single one, one obtains maximum AUC but zero precision.

This observation shows the need to go beyond single-layered scores and combine them into a more general metric that fully exploits the multiplex nature of the networks taken into account. Note, indeed, that single-layered metrics considered triadic relations among three nodes u, v, and w, in which the two links u-w and v-w both lie in the same layer. However, triadic relations in multiplex networks can be far richer [42,43]. Figure 1 shows different kinds of triadic relations in multiplex networks. Let us indicate as x the layer on which the link u-v is to be predicted. One can distinguish four types of triadic relations depending on the location of the (u, w) and (v, w) links: (i) $\mathcal{T}_{xx} = \{(u, v, w)|w \in \Gamma_x(u) \cap \Gamma_x(v)\}$, where both links lie in layer x; (ii) $\mathcal{T}_{x\alpha} = \{(u, v, w)|w \in \Gamma_x(u) \cap \Gamma_\alpha(v)\}$ and $\mathcal{T}_{\alpha x}$, where one link lies in the layer x and the other lies in another layer α; (iii) $\mathcal{T}_{\alpha\alpha} = \{(u, v, w)|w \in \Gamma_\alpha(u) \cap \Gamma_\alpha(v)\}$, where both links are in the same layer α, different from layer x; and (iv) $\mathcal{T}_{\alpha\beta} = \{(u, v, w)|w \in \Gamma_\alpha(u) \cap \Gamma_\beta(v)\}$ and $\mathcal{T}_{\beta\alpha}$, where one link is in layer α and the other in layer β, both different from layer x.
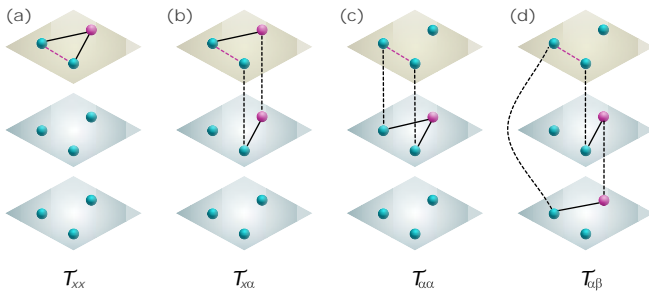

FIG. 1. Triadic relationships in a multiplex network. Given two nodes u and v for which we want to predict the future existence of a link (red dashed line) in the top layer x (green), based on their connections with another node w (pink) via triadic closure, we can distinguish four types of triadic relationships: (a) u and v are both connected to w in the prediction layer x; (b) the link between u and w is in the prediction layer x, but v and w are connected in a different layer α, or vice versa; (c) both u and v are connected to w in a layer α different from the prediction layer x; (d) u and w are connected in a layer α different from the prediction layer x, and v and w are connected in a third layer β different from layers α and x.

Within this formalism, one can consider a score that counts the common neighbors closing triads of each type, and weight each contribution by the logarithm of the degree, as in the Adamic-Adar score,

$$\text{MAA}(u, v) = \sum_{\alpha, \beta} \sum_{w \in \mathcal{T}_{\alpha\beta}} \frac{\eta_{x\alpha} \eta_{x\beta}}{\sqrt{\langle k \rangle_\alpha \langle k \rangle_\beta}} \frac{1}{\sqrt{\ln\left(k_w^\alpha\right) \ln\left(k_w^\beta\right)}}. \quad (2)$$

This expression is the generalization of the Adamic-Adar score for multiplex networks (MAA) with an arbitrary number of layers, in which the links to be predicted all lay in the same layer x. Several considerations are in order.

First, the contribution of each triad $(u, v, w) \in \mathcal{T}_{\alpha\beta}$ is weighted by the square root of the logarithm of the degree of node w in the two layers involving α and β. With this choice, the original weight $1/\ln(k_w)$ is naturally recovered for $\alpha = \beta = x$. Second, note that different layers of a multiplex network may show very different densities, as shown in Table S1 of the SM [41]. In case of similarity scores based on the number of common neighbors, as in this case, denser layers will have more triads and thus will be less informative. We take this into account by weighting the contribution of each type of triadic relation by the square root of the average degree of the layers involved, $\sqrt{\langle k \rangle_\alpha}$. Third, the coefficients $\eta_{x\alpha}$ before each term allow us to control the relative weight of each type of triadic closure in the total score of the link. We choose them in a way that $\eta_{x\alpha}$ corresponds to the weight of layer α. Without lack of generality, we choose $\sum_\alpha \eta_{x\alpha} = 1$. Fourth, the application of the AA score to layer α, corresponding to triad closures [Fig. 1(c)], is recovered by setting $\eta_{x\alpha} = 1$. The original AA score in single-layered networks [Fig. 1(a)] is recovered by simply setting $\eta_{xx} = 1$.

Figure 2 shows the AUC [Fig. 2(a)] and precision [Fig. 2(b)] of the MAA metric as a function of the coefficients $\eta_{x\alpha}$, for three of the eight data sets under consideration. Others are shown in Figs. S1 and S2 of the SM [41]. For the sake of convenient visualization, we consider only three layers for each network, to visualize the three coefficients in a triangle. For each network, we consider the prediction of links in each of the three layers. The coefficient $\eta_{x\alpha}$ indicates the weight of layer α in the prediction of new links in layer x. For instance, in the CNS data set, the coefficient η_{12} indicates the weight of Facebook friendship (represented in layer 2) in predicting new phone calls (layer 1). One can see that, in most cases, the maximum value of the AUC and the precision is achieved for nontrivial combinations of the coefficients, i.e., different from $\eta_{xx} = 1$ which corresponds to the classical AA score, showed in the left corner of triangles. This is particularly true for the precision, whose maximum is achieved in some cases in the middle of the triangle, i.e., with similar contributions for each layer, as in the case of the CKM or CNS networks. The exact values of the coefficients maximizing AUC and precision for each data set are reported in Table S2 of the SM [41].

Therefore, Fig. 2 shows that the prediction of a certain kind of links can be improved by exploiting additional, related information, encoded in other layers. For instance, Facebook friendship can help in predicting new calls [i.e., the maximum AUC for this task is obtained for $\eta_{12} = 0.40$—see plot (i) of Fig. 2(a) and Table S2 of the SM [41]], or additive genetic interactions and physical association can be predictive
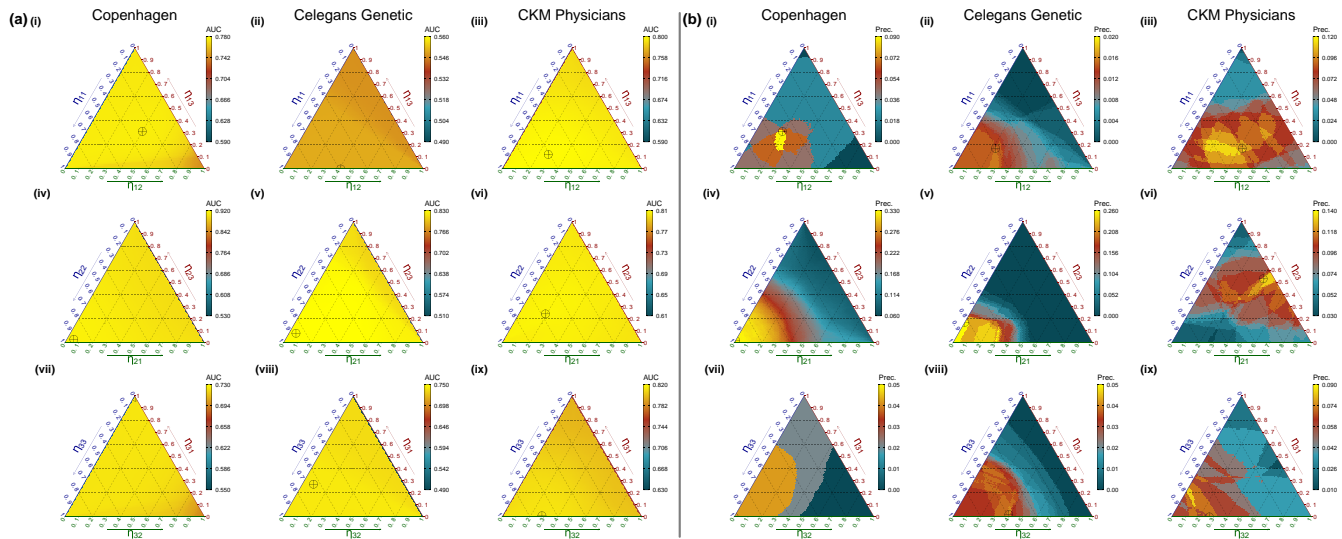
FIG. 2. (a) AUC and (b) precision of the MAA metric for different values of the coefficients $\eta_{x\alpha}$, indicating the weight of layer α in the prediction of new links in layer x. Varying the values of two coefficients, the third is naturally fixed. Each column corresponds to a different data set, represented as a multiplex network of three layers. Each row corresponds to a prediction in a different layer x (see Table S2 of the SM [41] for the corresponding interactions). A cross indicates the maximum value for each plot, corresponding to the combination of coefficients $(\eta_{x1}, \eta_{x2}, \eta_{x3})$, reported in Table S2 of the SM [41], that maximizes AUC or precision for the prediction of new links in layer x.

of direct protein interactions in *C. elegans* [i.e., the maximum precision is obtained for $\eta_{13} = 0.17$ and $\eta_{12} = 0.22$—see plot (ii) of Fig. 2(b) and Table S2 of the SM [41]]. Interestingly, this effect can be asymmetric: New offline interactions (calls and texts) are not predictive of Facebook friendships, as the corresponding coefficients η_{21} and η_{23} for this prediction task are zero. This is shown in plots (vii) of Figs. 2(a) and 2(b): The maximum value of the precision and AUC is obtained for $\eta_{22} \simeq 1$, in the left corner of the plots (see also Table S2 of the SM [41]). This implies that not all layers add valuable information for a specific link prediction task. In this case, a complete multiplex representation is redundant and such a layer can be effectively included in the others without missing relevant information.

Furthermore, we test if the MAA metric is able to optimally extract information from the multiplex representation, compared with the AA score applied to the aggregate network, that includes the same amount of information. Table II shows that the MAA metric outperforms the classical AA score with respect to both AUC and precision, in all data sets under consideration. Finally, in Table II we compare the MAA metric with other, state-of-the-art metrics for link prediction applied to the aggregated network representation, that includes all information available, in particular, common neighbors (CN), Jaccard's coefficient (JC), and preferential attachment (PA), which are based on the one-step neighborhoods of the nodes such as the AA score [44], and the Katz distance [45], which instead is based on path length. Table II shows the prediction of links in the first layer of each data set, and predictions in other layers are shown in Tables SIII and SIV of the SM [41], with similar results. One can see that the MAA metric outperforms the precision of all other metrics in all but one case (the Lazega data set), while it outperforms the AUC of other methods in five of eight data sets under consideration.

Before concluding, we stress that the metric encoded in Eq. (2) is different from previous extensions of link prediction to multilayer networks. Similarly, other approaches calculate the score of each layer and aggregate all of them (possibly with some weights), effectively neglecting structures of types $\mathcal{T}_{x\alpha}$ and $\mathcal{T}_{\alpha\beta}$ [25,46–48].

To sum up, we proposed a general method for link prediction that fully exploits different kinds of relational data encoded in several social and biological networks. Our metric is a generalization of the Adamic-Adar score for multiplex networks with an arbitrary number of layers, and it is able to

TABLE II. AUC (top) and precision (bottom) to predict the first layer of each data set, obtained for different metrics: Adamic-Adar (AA), common neighbors (CN), Jaccard similarity (JC), Katz distance (Katz), preferential attachment (PA), and multiplex Adamic-Adar (MAA). All metrics except for the MAA are applied to the aggregated network, including all information available. The best method is highlighted in bold.

Method	CNS	CEG	CEN	CSA	CKM	EUA	VIC	LAZ
AA	0.76	0.55	0.79	0.79	0.80	0.86	0.68	0.66
CN	0.76	0.55	0.78	0.78	0.80	0.86	0.67	0.66
JC	0.76	0.55	0.77	0.88	0.80	0.82	0.68	0.70
Katz	0.72	0.59	0.77	0.74	**0.91**	0.89	0.65	0.58
PA	0.58	**0.65**	0.62	0.39	0.64	**0.90**	0.61	0.54
MAA	**0.77**	0.55	**0.79**	**0.91**	0.80	0.87	**0.71**	**0.71**
AA	0.00	0.00	0.03	0.03	0.09	0.00	0.35	0.14
CN	0.00	0.00	0.03	0.03	0.10	0.00	0.35	0.13
JC	0.00	0.00	0.02	0.32	0.08	0.00	0.40	**0.23**
Katz	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.00
PA	0.00	0.00	0.02	0.00	0.01	0.02	0.27	0.09
MAA	**0.09**	**0.02**	**0.08**	**0.39**	**0.11**	**0.19**	**0.46**	0.21

outperform single-layered AA scores in all considered data sets. The MAA metric also outperforms several well-known link prediction algorithms, such as the Jaccard's coefficient or the Katz distance. The coefficients $\eta_{x\alpha}$ that maximize the MAA score have an interesting interpretation, as they correspond to the weight to be assigned to each layer in order to optimize the information structured in the network for the link prediction task, indicating which layers are redundant. Interestingly, this effect can be asymmetric with respect to predictions in different layers. The computational complexity of the MAA metric is similar to other similarity-based scores. With respect to the classical AA score, it increases with the number of layers in the multiplex network, which is usually small. Note that the triadic relationships need to be computed just once and stored, then the whole range of coefficients

can be scanned to obtain the ones that maximize the MAA score. In future works, it would be interesting to generalize to multiplex networks other metrics based on single layers, such as the Katz distance, which is based on paths that can be reconstructed across layers.

[1] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

[2] B. Taskar, M. Fai Wong, P. Abbeel, and D. Koller, in *Advances in Neural Information Processing Systems 16*, edited by S. Thrun, L. K. Saul, and B. Schölkopf (MIT Press, Cambridge, MA, 2004), pp. 659–666.

[3] L. Lu and T. Zhou, Physica A **390**, 1150 (2011).

[4] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10* (ACM Press, New York, 2010), pp. 243–252.

[5] L. A. N. Amaral, Proc. Natl. Acad. Sci. USA **105**, 6795 (2008).

[6] D. Liben-Nowell and J. Kleinberg, in *Proceedings of the Twelfth International Conference on Information and Knowledge Management, CIKM '03* (ACM Press, New York, 2003), pp. 556–559.

[7] M. A. Hasan and M. J. Zaki, in *Social Network Data Analytics*, edited by C. C. Aggarwal (Springer, Boston, MA, 2011), pp. 243–275.

[8] B. Taskar, P. Abbeel, and D. Koller, in *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, UAI'02* (Morgan Kaufmann, San Francisco, CA, 2002), pp. 485–492.

[9] W. Liu and L. Lu, Europhys. Lett. **89**, 58007 (2010).

[10] R. Guimerà and M. Sales-Pardo, Proc. Natl. Acad. Sci. USA **106**, 22073 (2009).

[11] M. Marjan, N. Zaki, and E. A. Mohamed, in *Proceedings of the 2018 IEEE 5th International Congress on Information Science and Technology (CiSt)* (IEEE, New York, 2018), p. 200.

[12] V. Martínez, F. Berzal, and J. C. Cubero, ACM Comput. Surv. **49**, 69 (2016).

[13] C. A. Bliss, M. R. Frank, C. M. Danforth, and P. S. Dodds, J. Comput. Sci. **5**, 750 (2014).

[14] L. A. Adamic and E. Adar, Soc. Netw. **25**, 211 (2003).

[15] A. Clauset, C. Moore, and M. E. J. Newman, Nature (London) **453**, 98 (2008).

[16] T. Jia, Y. Ran, and X. Xu, in *Proceedings of NetSci-X 2020: Sixth International Winter School and Conference on Network Science* (Springer, Berlin, 2020).

[17] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, J. Complex Netw. **2**, 203 (2014).

[18] A. Aleta and Y. Moreno, Annu. Rev. Condens. Matter Phys. **10**, 45 (2019).

[19] D. Davis, R. Lichtenwalter, and N. V. Chawla, Soc. Netw. Anal. Min. **3**, 127 (2013).

[20] K.-K. Kleineberg, M. Boguñá, M. Ángeles Serrano, and F. Papadopoulos, Nat. Phys. **12**, 1076 (2016).

[21] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, in *Proceedings of the SDM06: Workshop on Link Analysis, Counterterrorism and Security* (SIAM, Maryland, 2006), Vol. 30, pp. 798–805.

[22] Z. Lu, B. Savas, W. Tang, and I. S. Dhillon, in *Proceedings of the 2010 IEEE International Conference on Data Mining* (IEEE, New York, 2010), p. 923.

[23] R. Matsuno and T. Murata, in *Companion Proceedings of the The Web Conference 2018, WWW '18* (International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018), pp. 1261–1268.

[24] M. Pujari and R. Kanawati, Netw. Heterog. Media **10**, 17 (2015).

[25] K. Sotiropoulos, J. W. Byers, P. Pratikakis, and C. E. Tsourakakis, in *Proceedings of the 57th Annual Allerton Conference on Communication, Control, and Computing* (IEEE, New York, 2019), pp. 973–980.

[26] A. Hajibagheri, G. Sukthankar, and K. Lakkaraju, in *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE, New York, 2016), p. 1079.

[27] L. Yao, L. Wang, L. Pan, and K. Yao, Proc. Comput. Sci. **83**, 82 (2016).

[28] M. Jalili, Y. Orouskhani, M. Asgari, N. Alipourfard, and M. Perc, R. Soc. Open Sci. **4**, 160863 (2016).

[29] D. Hristova, A. Noulas, C. Brown, M. Musolesi, and C. Mascolo, EPJ Data Sci. **5**, 1 (2016).

[30] H. Mandal, M. Mirchev, S. Gramatikov, and I. Mishkovski, in *Proceedings of the 2018 26th Telecommunications Forum (TELFOR)* (IEEE, New York, 2018), p. 1.

[31] R. R. Junuthula, K. S. Xu, and V. K. Devabhaktuni, arXiv:1804.08584.

[32] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, Nat. Commun. **6**, 6864 (2015).

[33] P. Sapiezynski, A. Stopczynski, D. D. Lassen, and S. Lehmann, Sci. Data **6**, 315 (2019).

[34] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, Nucl. Acids Res. **34**, D535 (2006).

[35] B. L. Chen, D. H. Hall, and D. B. Chklovskii, Proc. Natl. Acad. Sci. USA **103**, 4723 (2006).

[36] M. Magnani, B. Micenkova, and L. Rossi, arXiv:1303.4986.

[37] J. Coleman, E. Katz, and H. Menzel, Sociometry **20**, 253 (1957).

[38] A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, and S. Boccaletti, Sci. Rep. **3**, 1344 (2013).

[39] E. Lazega, *The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership* (Oxford University Press, Oxford, UK, 2001).

[40] M. Vickers and S. Chan, *Representing Classroom Social Structure* (Victoria Institute of Secondary Education, Melbourne, 1981).

[41] See Supplemental Material at http://link.aps.org/supplemental/ 10.1103/PhysRevResearch.2.042029 for Figures, and Tables.

[42] F. Battiston, V. Nicosia, and V. Latora, Phys. Rev. E **89**, 032804 (2014).

[43] E. Cozzo, M. Kivelä, M. De Domenico, A. Solé-Ribalta, A. Arenas, S. Gómez, M. A. Porter, and Y. Moreno, New J. Phys. **17**, 073029 (2015).

[44] F. Gao, K. Musial, C. Cooper, and S. Tsoka, Sci. Program. **2015**, 172879 (2015).

[45] L. Katz, Psychometrika **18**, 39 (1953).

[46] S. Sharma and A. Singh, *An Efficient Method for Link Prediction in Complex Multiplex Networks* (IEEE, New York, 2015).

[47] Y. Yao, R. Zhang, F. Yang, Y. Yuan, Q. Sun, Y. Qiu, and R. Hu, Int. J. Mod. Phys. C **28**, 1750101 (2017).

[48] Z. Samei and M. Jalili, J. Complex Netw. **7**, 641 (2019).