Vector Centrality in Networks with Higher-Order Interactions

K. Kovalenko,1,2 M. Romance,2 D. Aleja,2 E. Vasilyeva,1,3 R. Criado,2 A.M. Raigorodskii,1,4,5,6 J. Flores,2 K. Alfaro-Bittner,7 M. Perc,8,9,10,11 and S. Boccaletti1,2,12

1Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny, 141701, Russia
2Universidad Rey Juan Carlos, Calle Tulipán sn, 28933 Móstoles, Madrid, Spain
3P.N. Lebedev Physical Institute of the Russian Academy of Sciences, 53 Leninskiy prosp., 119991 Moscow, Russia
4Caucasus Mathematical Center at Adyghe State University, ul. Pervomaiskaya, 208, Maykop, 385000, Russia
5Mechanics and Mathematics Faculty, Moscow State University, Leninskie Gory, 1, Moscow, 119991, Russia
6Institute of Mathematics and Computer Science, Buryat State University, ul. Ranzhurova, 5, Ulan-Ude, 670000, Russia
7Department of Física, Universidad Técnica Federico Santa María, Av. España 1680, Casilla 110V, Valparaíso, Chile
8Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia
9Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404332, Taiwan
10Complexity Science Hub Vienna, Josefstaedterstraße 39, 1080 Vienna, Austria
11Alma Mater Europaea, Slovenska ulica 17, 2000 Maribor, Slovenia
12CNR - Institute of Complex Systems, Via Madonna del Piano 10, I-50019 Sesto Fiorentino, Italy

(Dated: September 1, 2021)

Identifying the most influential nodes in networked systems is vital to optimize their function and control. Several scalar metrics have been proposed to that effect, but the recent shift in focus towards higher-order networks has rendered them void of performance guarantees. We propose a new measure of node’s centrality, which is no longer a scalar value, but a vector with dimension one lower than the highest order of interaction in the graph. Such a vectorial measure is linked to the eigenvector centrality for networks containing only pairwise interactions, whereas it has a significant added value in all other situations where interactions occur at higher-orders. In particular, it is able to unveil different roles which may be played by a same node at different orders of interactions, an information which is impossible to be retrieved by single scalar measures.

Ranking nodes in a graph is the most fundamental task in modern network science [1–4]. Already in 1977, Linton C. Freeman gave the first definition of betweenness centrality, and used it to rank individual clout in social networks [5–6]. The earliest definition and use of eigenvector centrality can even be traced more than a century ago, in 1895, when Edmund Landau used it for scoring chess tournaments [7]. Nonetheless, it was not before the discovery of heterogeneity in the degree distributions of real world networks [8] that the full depth of implications of node centrality was realized. The ‘hub’ became, and still is, a popular meme that stands for influence, importance, or virality in social, biological and technological networks [1,9,12]. And the identification of the most central nodes in complex networks is crucial for error and attack tolerance [13,14], viral marketing [15], information spreading [16–18], influence maximization [19,20], as well as optimal immunization against epidemics [21,25], plant genomic [26] and cancer research [27,28], just to name a few examples. Not to mention that companies like Google are actually building their all business in providing efficient and customized rankings of webpages.

Although the paramounct of quantifying node centrality is undisputed, the best measure for it very much depends on the particularities of the problem at hand. The various measures adopted so far to quantify node centrality, from the simplest node degree to the variations of betweenness and eigenvector centrality [29–32], do not optimize a global function of influence, and are thus inherently unable to guarantee optimal performance [19]. Therefore, the correct question one has to ask himself is not how central is a given node in a network, but rather how central is a given node in a network with respect to a given process. The issue is further exacerbated by the recent departures from traditional networks towards multilayer and higher-order networks as more apt representations of real world systems [33–35]. Although a generalization of eigenvector centrality for multiplex networks has been proposed [36], this does not account for the fact that in higher-order networks a link can connect more than two nodes. The potential of high-order interactions has been recognized already in the early 70s by Ronald H. Atkin [37], but the interest peaked only recently with the inability of classic graph representations to describe group interactions. This ineptitude comes to a head when studying peer pressure, public cooperation, complex contagion, or opinion formation, to list just a few examples that clearly extend well beyond pairwise interactions in social science [38,39], or when considering three or more species that routinely compete for food and territory in a complex ecosystem [40], or when functional [41] or structural [42] brain networks or protein interaction networks [43] are studied. In view of these recent developments, it is therefore crucial and of wide interest to generalize centrality measures to account for high-order interactions.

In this Letter, we consider the most general case of an ensemble of $N$ nodes which interplay by means of interactions of any order $d < D$ (with $D$ indicating the maximum order of group interactions taking place in the ensemble), and introduce a novel measure of centrality which is a vector assigned to each node, with dimension $D - 1$. While our vector centrality is related to the classical eigenvector centrality for networks containing only pairwise interactions, we will demonstrate that our new measure has instead a significant added value in all situations where interactions occur at
higher-orders. We will show with three practical applications that our measure is, indeed, able to distinguish different roles which may be played by the same node at different orders of interactions, a feature which is evidently impossible to be revealed by any single scalar measure.

Let us start then with considering $N$ nodes which are interplaying by means of $l_2$ links (pairwise interactions), $l_3$ hyperlinks of order 3 (triadwise interactions), $l_4$ hyperlinks of order 4 (quadwise interactions), and in general by $l_d$ hyperlinks of order $d$ (with $d = 2, 3, ..., D$). We here concentrate on the case where all such hyperlinks are undirected. Mathematically this defines an undirected higher-order network (or hypergraph) $G = (V, E)$, i.e. a finite set $V$ containing $N$ nodes, and a family $E$ of $l_N = \sum_{d=2}^{D} l_d$, non-empty subsets of nodes of $G$, each subset defining a hyperlink.

Our idea is to associate to $G$ its linegraph $L(G)$, as introduced by Hassler Whitney for graphs in 1932 [44] and extended for higher-order networks by Jean-Claude Bermond et al. in 1977 [45, 46]. In particular, $L(G)$ is a graph of $l_N$ nodes (each of which mapping one of the hyperedges of $G$). The links of $L(G)$ stand for adjacency between hyperedges in $G$: if $h_1 \in E$ and $h_2 \in E$ are two hyperlinks, then there is an undirected link in $L(G)$ between the nodes $h_1$ and $h_2$ if and only if $h_1 \cap h_2 \neq \emptyset$. Figure 1 depicts an illustrative example, where a hypergraph $G = (V, E)$ is defined by $V = \{1, 2, 3, 4, 5\}$ and $\mathcal{E} = \{\{1, 2, 3\}, \{1, 2, 3, 4\}, \{2, 4, 5\}, \{4, 5\}\}$. The Figure shows also the associated linegraph $L(G)$, and the projection network $\pi_2(G)$ of $G$, a classic network which is formed by the same set of nodes as in $G$ and whose links represent the pairwise interactions resulting from the projection of the hyperlinks of $G$.

Now, it is straightforward to demonstrate that if $G$ is undirected and connected, then also $L(G)$ is undirected and connected. Indeed, for any pair of hyperedges $h_i$ and $h_j$ in $L(G)$ a path can be constructed by choosing a node $v$ from $h_i$ and a node $w$ from $h_j$, and by using the same sequence of hyperedges as in path from $v$ to $w$ in $G$. Then, the classic Perron-Frobenius theorem [47, 48] guarantees the existence and uniqueness of the eigenvector centrality of $L(G)$. In other words, one can compute with standard methods the classical eigenvector centrality of each node on $L(G)$, and one obtains a value $c(h) \in [0, 1]$ for all hyperlinks $h \in \mathcal{E}$ in $G$, such that $\sum_{h \in \mathcal{E}} c(h) = 1$.

With the $l_N$ values of $c(h)$ at hand, we can now define the vector centrality of each node $i \in V$, a non-negative vector $\vec{c}_i = (c_{i1}, \cdots, c_{iD}) \in \mathbb{R}^{D-1}$ such that, for every $2 \leq k \leq D$ one has

$$c_{ik} = \frac{1}{k} \sum_{|h| = k} \sum_{i \in h} c(h),$$

where $|h|$ indicates the length (or size) of the hyperedge $h$, and $D = \max \{|h|; \ h \in \mathcal{E}\}$ is the maximal size of hyperedges in $G$ (the maximal order of the group interactions affecting the $N$ nodes in the ensemble). In other words, the $k^{th}$ component $c_{ik}$ of the vector centrality of node $i$ is the sum of the centralities of all hyperlinks of size $k$ that contain $i$ as one of the incident nodes, and the weight value $\frac{1}{k}$ makes that

$$\sum_{i \in V} \frac{\|i\|_1}{\|i\|} = \sum_{i \in V} \sum_{k=1}^{D} c_{ik} = \sum_{i \in V} \sum_{k=1}^{D} \sum_{|h| = k} c(h) \frac{1}{k} = \sum_{h \in \mathcal{E}} \sum_{i \in h} \frac{c(h)}{|h|} = \sum_{h \in \mathcal{E}} \frac{c(h)}{|h|} = \sum_{h \in \mathcal{E}} c(h) = 1,$$

since every summand $\frac{c(h)}{|h|}$ appears exactly $|h|$ times and therefore it is normalised.

Notice that if $D = 2$, i.e., only pairwise interactions exist in $G$, then for each node $i$ one has $c_i = (c_{i2}) \in \mathbb{R}$, where the scalar value $c_{i2}$ is related with the $i^{th}$ component ($c_{i2}$) of the classic eigenvector centrality of $G$, as it was proved in [49]. Precisely, for $D = 2$, calling $\lambda_1$ and $\lambda_2$ the greatest and second greatest eigenvalue of the adjacency matrix of $G$, and denoting by $\Delta$ the norm of the difference between our measure and the eigenvector centrality, Ref. [50] gave the following bounding relationship:

$$\Delta \equiv \sqrt{\frac{1}{N} \sum_{i=1}^{N} (c_{i2} - c'_i)^2} \leq \frac{4 - \sqrt{2} \sqrt{2N \sqrt{2D_2}}}{\lambda_1 - \lambda_2} \sqrt{\lambda_1 - \frac{2\lambda_2}{N}},$$
The first example refers to a synthetic hypergraph consisting of \( N = 100 \) nodes, and \( l_2 = l_3 = l_4 = 400 \) hyperlinks, mapping therefore an ensemble of units interplaying by means of pairwise, triadwise and quadwise interactions. Here, we want to highlight how our vectorial centrality outperforms classical measures in tracking the importance of nodes when changes occur in the network structure. To this purpose, we initially prepare a graph with all \( l_3 \) hyperlinks which are randomly distributed. As instead for the \( l_2 \) links (the \( l_4 \) hyperlinks), 350 of them are distributed randomly, whereas 50 of them are placed so as to make vertex 1 (vertex 100) a hub for pairwise (quadwise) interactions, i.e., they are constructed so as to include vertex 1 (vertex 100) as one of the incident nodes. Then, we simulate limitations processes in group interactions (like for instance those capacity restrictions in public places enforced by Governments to maintain social distancing measures during the containment of the COVID19 health emergency) by removing at random a fraction \( p \) of quadwise interactions, and we keep track on how the different centrality measures are efficient in monitoring the change of relevance of each node following the changes in the network structure. Precisely, one surveils the behavior of \( c_{1,2} \) (the first component of the vector centrality of node 1), of \( c_{N,4} \) (the last component of the vector centrality of node 100), \( c_{1,\pi} \) and \( c_{N,\pi} \) (the first and last components of the classical eigenvector centrality for the projected graph, \( \vec{c}_1 \)), and \( c_{1,\pi_\omega} \) and \( c_{N,\pi_\omega} \) (the first and last components of the classical eigenvector centrality for the projected weighted graph, \( \vec{c}_{\pi_\omega} \)). In \( \pi \) two nodes are connected if there exists a hyper-link to which they both belong to; in \( \pi_\omega \), the weight of each link is the number of hyperlinks to which the two end nodes are belonging to.

The results are reported in panel (a) of Fig. 2, and show clearly that only our vectorial measure (by comparison of \( c_{1,2} \) and \( c_{N,4} \)) is able to reveal a substantial loss of centrality of node 100 as the number of quadwise interactions is progressively reduced, and a corresponding gain in centrality of node 2 which eventually remains the only hub in the system.

In the second example, we probe the capability of our measure to reveal different scaling properties which may affect different orders of interactions in the graph, even in the case in which, at variance with the previous example, such orders do not correspond to the same number of hyperlinks. To that purpose, we construct another synthetic graph with pairwise, triadwise and quadwise interactions, this time with \( N = 1,000 \), \( l_2 = 4,000 \), \( l_3 = 1,000 \) and \( l_4 = 2,000 \). All \( l_2 \) and \( l_4 \) hyperlinks are chosen randomly, this way determining a strongly homogeneous distribution for pairwise and quadwise interactions. Instead, the \( l_3 \) hyperlinks (which are in the minimum number with respect to all other hyperlinks) are chosen so as to determine a strongly heterogeneous distribution: at each time those hyperlinks are constructed with a probability which explicitly depends on the actual node degree.

It has to be remarked that any projection (weighted or unweighted) of such synthetic hypernetwork would result in a heterogeneous degree distribution, with the consequence that any centrality measure applied to such projected graph would reveal a strong heterogeneity. The results of applying our measure are, instead, shown in panels (b-d) of Fig. 2 where we report the histograms of the first (\( c_2 \), panel b), the second
(c_3, panel c) and the third (c_4, panel d) component of our vectorial centrality. It is seen that while the histograms reveal strong homogeneity at the level of pairwise and quadrwise interactions (panels b and d), they clearly show heterogeneity traits at the level of triadwise interactions (panel c), this way accounting exhaustively for the overall structural properties which have been engineered in the hypergraph.

Finally, the third example refers to a real-world hypergraph, mapping the information publicly available in the arXiv (https://arxiv.org/ https://github.com/mattbierbaum/arxiv-public-datasets/) database, with the data parsing made by Ref. [51]. In particular, we focused on the data of preprints published in mathematics, and extracted those papers which were written in collaboration, i.e. those having at least two co-Author. The extracted dataset consists of a total of 498,071 papers co-written by 230,605 Authors.

The data were mapped into a hypergraph $G_{math}$, where nodes were scientists, and each paper formed a hyperlink (a group interaction) of length equal to the number of co-Author. The maximal number of co-Author of a single papers (i.e. the maximal length $D$ of hyperlinks in $G_{math}$) is 67, which implies that the vectorial centrality of each scientist will have 66 components. The associated linegraph $L(G_{math})$ is rather large in size: it is obviously formed by 498,071 nodes, and it has $9,808,188$ links. The eigenvector centrality of $L(G_{math})$ is then calculated, and the vector centrality of each scientist in $G_{math}$ is evaluated.

Various rankings of scientists may be extracted according to the different components of the vector centrality, i.e. scientists may have different role and importance with respect to different hyperedges’ sizes. In particular, we here analyze how many of the members of the top $x$ authors’s list in the ranking with respect to a given component of the vector centrality is also belonging to the top $x$ authors’s list in the ranking made with respect to another component. To do so, we introduce the fraction $\mu_x$ as follows:

$$\mu_x(c^j, c^l) = \frac{|\text{top}_x(c^j) \cap \text{top}_x(c^l)|}{x},$$

(2)

where $c^j$ and $c^l$ are, respectively, the $i^{th}$ and $j^{th}$ components of the vector centrality of the nodes, $\text{top}_x(c^j)$ ($\text{top}_x(c^l)$) is the set of the nodes which are occupying the top $x$ positions in the ranking made by comparing the $i^{th}$ (the $j^{th}$) component of their vector centralities, and $|\cdot|$ stands here for the cardinality of the set. $\mu_x(c^j, c^l)$ measures therefore how large is the overlap between the two sets, and its values $\mu_x(c^j, c^l)$ form a square matrix of $66 \times 66$ elements, which actually describes how correlated are the positions scientists are holding in the ranking calculated with respect to a given component of the vector centrality with those held by the same scientists in the ranking calculated with respect to another component.

The values of $\mu_{100}(c^i, c^j)$ (limited to the first ten components, out of the 66, of the vector centrality) are reported in Fig. [5]. It is clearly seen that, except for the few values close to $i = j$, the fractions $\mu_{100}(c^i, c^j)$ are relatively small for $i \neq j$ and, therefore, the lists of the 100 top leaders in the rankings made with respect to different hyperedges sizes are significantly different. This confirms that the use of our measure is essential for extracting information on such differences, which would be instead unaccessible by any other scalar measure of centrality.

Taken together, we have introduced a centrality measure in order to overcome the inherent limitations of scalar centralities in higher-order networks. Our measure assigns a vector to each node, with dimension one lower than the dimension of the longest hyperlink in the network, and with every component thus determining the centrality of that node in a link with a particular length. By using artificially generated higher-order networks as well as data from the arXiv mathematics section, we further demonstrated that our measure is able to unveil different roles which may be played by a same node at different orders of interactions, and therefore it is the only one which accounts exhaustively for the properties of the overall interactive structure of the hypergraph.

As noted already when introducing our vector centrality, the same approach can be readily applied to other structural measures, which thus opens the path towards a wider applicability of our approach. From finding the most influential nodes for an optimal design of strategies that arrest an epidemic to the maximization of information flow or influence, we expect our measure to become widely used with further progress in network science and related research fields.

This work was supported by the Russian Federation Government (Grant No. 075-15-2019-1926 and project "Post-crisis world order: challenges and technologies, competition and cooperation" funded by the Ministry of Science and Higher Education, agreement number 075-15-2020-783), by the program "Leading Scientific Schools " (Grant No. NSh-2540.2020.1), by the Spanish Government (Project PGC2018-101625-B-I00 (AEI/FEDER, UE)), by the URJC Grant No.
M1993, and by the Slovenian Research Agency (Grants No. P1-0403 and J1-2457).

* The two authors have contributed equally to the Manuscript
† These two authors equally contributed to the Manuscript

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, Phys. Rep. 424, 175 (2006).
[2] M. E. J. Newman, Networks (Oxford University Press, Oxford, U.K., 2010).
[3] M. O. Jackson, Social and Economic Networks (Princeton University Press, Princeton, NJ, 2010).
[4] A.-L. Barabási, Network Science (Cambridge University Press, Cambridge, U.K., 2016).
[5] L. C. Freeman, Sociometry 40, 35 (1977).
[6] L. C. Freeman, Soc. Netw. 1, 215 (1978).
[7] E. Landau, Deutsches Wochenschach 11, 3 (1895).
[8] A.-L. Barabási and R. Albert, Science 286, 509 (1999).
[9] R. Albert and A.-L. Barabási, Rev. Mod. Phys. 74, 47 (2002).
[10] M. E. J. Newman, SIAM Review 45, 167 (2003).
[11] M. Barthélemy, Phys. Rep. 499, 1 (2011).
[12] P. Holme and J. Saramäki, Phys. Rep. 519, 97 (2012).
[13] R. Albert, H. Jeong, and A.-L. Barabási, Nature 406, 378 (2000).
[14] R. Cohen, K. Erez, D. Ben-Aviraham, and S. Havlin, Phys. Rev. Lett. 86, 3682 (2001).
[15] M. Richardson and P. Domingos, in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (2002), pp. 61–70.
[16] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, Nat. Phys. 6, 888 (2010).
[17] J. M. Kleinberg, 46, 604 (1999).
[18] M. P. van den Heuvel and O. Sporns, Trends Cogn. Sci. 17, 683 (2013).
[19] F. Morone and H. A. Makse, Nature 524, 65 (2015).
[20] L. Li, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, Phys. Rep. 650, 1 (2016).
[21] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. 86, 3200 (2001).
[22] M. E. Newman, Phys. Rev. E 66, 016128 (2002).
[23] Y. Chen, G. Paul, S. Havlin, F. Liljeros, and H. E. Stanley, Phys. Rev. Lett. 101, 058701 (2008).
[24] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Rev. Mod. Phys. 87, 925 (2015).
[25] Z. Wang, C. T. Bauch, S. Bhattacharyya, A. d’Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, and D. Zhao, Phys. Rep. 664, 1 (2016).
[26] M. Zanin, J. Medina Alcazar, J. V. Carbajosa, M. Gomez Paez, D. Papo, P. Sousa, E. Menalvas, and S. Boccaletti, Sci. Rep. 4, 5112 (2014).
[27] A. Karsakov, T. Bartlett, A. Ryblov, I. Meyerov, M. Ivanchenko, and A. Zaikin, PLOS ONE 12, 1 (2017).
[28] H. J. Whitwell, O. Blyuss, U. Menon, J. F. Timms, and A. Zaikin, Oncotarget 9, 22717 (2018).
[29] U. Brandes, J. Math. Soc. 25, 163 (2001).
[30] M. E. Newman, Soc. Netw. 27, 39 (2005).
[31] M. Piraveenan, M. Prokopenko, and L. Hossain, PLOS One 8, e53095 (2013).
[32] A. Alvarez-Socorro, G. Herrera-Almarza, and L. González-Díaz, Sci. Rep. 5, 17095 (2015).
[33] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, Phys. Rep. 544, 1 (2014).
[34] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, J. Complex Netw. 2, 203 (2014).
[35] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petro, Phys. Rep. 874, 1 (2020).
[36] L. Solá, M. Romance, R. Criado, J. Flores, A. García del Amo, and S. Boccaletti, Chaos 23, 033131 (2013).
[37] R. H. Atkin, Int. J. Man-Mach. Stud. 4, 139 (1972).
[38] G. Burgio, J. T. Matamalas, S. Gómez, and A. Arenas, Entropy 22, 744 (2020).
[39] U. Alvarez-Rodriguez, F. Battiston, G. F. de Arruda, Y. Moreno, M. Perc, and V. Latora, Nat. Hum. Behav. 5, 586 (2021).
[40] J. M. Levine, J. Bascompte, P. B. Adler, and S. Allesina, Nature 546, 56 (2017).
[41] L.-D. Lord, P. Expert, H. M. Fernandes, G. Petri, T. J. Van Hartenvel, F. Vaccarino, G. Deco, F. Turkheimer, and M. L. Kringlebach, Front. Syst. Neurosci. 10, 85 (2016).
[42] A. E. Sizemore, C. Giusti, A. Kahn, J. M. Vettel, R. F. Betzel, and D. S. Bassett, J. Comput. Neurosci. 44, 115 (2018).
[43] E. Estrada and G. J. Ross, J. Theor. Biol. 438, 46 (2018).
[44] H. Whitney, Am. J. Math. 54, 150 (1932).
[45] J.-C. Bermond, M.-C. Heydemann, and D. Sotteau, Discrete Math. 18, 235 (1977).
[46] M.-C. Heydemann and D. Sotteau, Colloq. Math. Soc. J. Bolyai 18, 567 (1976).
[47] O. Perron, Mathematische Annalen 64, 248 (1907).
[48] F. Frobenius, Sitzungsberichte Akad. Wiss. Berlin 26, 456 (1912).
[49] R. Criado, J. Flores, A. García del Amo, and M. Romance, J. Comput. Appl. Math. 235, 1775 (2011).
[50] R. Criado, J. Flores, A. García del Amo, and M. Romance, Int. J. Comput. Math. 91, 304 (2014).
[51] C. B. Clement, M. Bierbaum, K. P. O’Keeffe, and A. A. Alemi, arXiv 19, 1905.00075 (2019).