

# SCIENTIFIC REPORTS



OPEN

## Distinct types of eigenvector localization in networks

Romualdo Pastor-Satorras<sup>1</sup> & Claudio Castellano<sup>2,3</sup>

Received: 21 April 2015

Accepted: 27 November 2015

Published: 12 January 2016

The spectral properties of the adjacency matrix provide a trove of information about the structure and function of complex networks. In particular, the largest eigenvalue and its associated principal eigenvector are crucial in the understanding of nodes' centrality and the unfolding of dynamical processes. Here we show that two distinct types of localization of the principal eigenvector may occur in heterogeneous networks. For synthetic networks with degree distribution  $P(q) \sim q^{-\gamma}$ , localization occurs on the largest hub if  $\gamma > 5/2$ ; for  $\gamma < 5/2$  a new type of localization arises on a mesoscopic subgraph associated with the shell with the largest index in the  $K$ -core decomposition. Similar evidence for the existence of distinct localization modes is found in the analysis of real-world networks. Our results open a new perspective on dynamical processes on networks and on a recently proposed alternative measure of node centrality based on the non-backtracking matrix.

An issue of paramount significance regarding the analysis of networked systems is the identification of the most important (or *central*) vertices<sup>1</sup>. The *centrality* of a vertex may stem from the number of different vertices that can be reached from it, from the role it plays in the communication between different parts of the network, or from how closely knit its neighborhood is. Following these approaches, different centrality measures have been defined and exploited, such as degree centrality, betweenness centrality<sup>2</sup>, or the  $K$ -core index and associated  $K$ -core decomposition<sup>3</sup>. Among those definitions, one of the most relevant is based on the intuitive notion that nodes are central when they are connected to other central nodes. This concept is mathematically encoded in the *eigenvector centrality*<sup>4</sup> (EC) of node  $i$ , defined as the component  $f_i$  of the principal eigenvector (PEV)  $\mathbf{f}$  associated with the largest eigenvalue  $\Lambda_1$  of the adjacency matrix  $A_{ij}$ . EC is the simplest of a family of centralities based on the spectral properties of the adjacency matrix including, among others, Katz's centrality<sup>5</sup> and PageRank<sup>6</sup>.

Apart from providing relevant information about the network structure<sup>1</sup>, the PEV and associated largest eigenvalue play a fundamental role in the theoretical understanding of the behavior of dynamical processes, such as synchronization<sup>7</sup> and spreading<sup>8,9</sup>, mediated by complex topologies. Considerable effort has thus been devoted in recent years to the study of the spectral properties of heterogeneous networks<sup>10–13</sup>. In this framework, Goltsev *et al.*<sup>14</sup> (see also<sup>15,16</sup>) have considered the *localization* of the PEV, i.e., whether its normalization weight is concentrated on a small subset of nodes or not. More in detail, let us consider an ensemble of networks of size  $N$ , with a PEV  $f_i$  normalized as a standard Euclidean vector, i.e.  $\sum_i f_i^2 = 1$ . An eigenvector is *localized* on a subset  $V$  of size  $N_V$  if a finite fraction of the normalization weight is concentrated on  $V$  ( $\sum_{i \in V} f_i^2 \sim \mathcal{O}(1)$ ) despite the fact that  $V$  is not extensive, i.e.,  $N_V$  is not proportional to  $N$ . This includes the case of localization on a finite set of nodes (i.e.  $N_V$  independent of  $N$ ,  $N_V = 1$  in the extreme case of localization on a single node), but also the case of localization on a mesoscopic subset of nodes for which  $N_V \sim N^\beta$  with  $\beta < 1$ . Otherwise, the eigenvector is instead *delocalized*, and a finite fraction of the nodes  $N_V \sim N$  contribute to the normalization weight, implying that their components are  $f_i \sim N^{-1/2}$ .

In this context, Goltsev *et al.*<sup>14</sup> study the localization in power-law distributed networks, with a degree distribution scaling as  $P(q) \sim q^{-\gamma}$ , for which the leading eigenvalue  $\Lambda_1$  is essentially given by the maximum between  $\langle q^2 \rangle / \langle q \rangle$  and  $\sqrt{q_{\max}}$ , where  $q_{\max}$  is the largest degree in the network<sup>9,11</sup>. For  $\gamma > 5/2$ , where  $\Lambda_1 \sim \sqrt{q_{\max}}$ , Goltsev *et al.*<sup>14</sup> find that the PEV becomes localized around the hub with degree  $q_{\max}$ <sup>14</sup>. On the other hand, they argue that, for  $\gamma < 5/2$ , when  $\Lambda_1 \sim \langle q^2 \rangle / \langle q \rangle$ , the PEV is delocalized. These observations are relevant in different contexts. Firstly, they point out a weakness of EC as a measure of centrality for heterogeneous power-law networks ( $\gamma > 5/2$ ), because of the exceedingly large role of the largest hub<sup>16</sup>. On the other hand, in the so-called quenched mean-field approach<sup>9,17</sup> to epidemic spreading on networks, the density of infected individuals in the steady state can be related

<sup>1</sup>Departament de Física, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain. <sup>2</sup>Istituto dei Sistemi Complessi (ISC-CNR), Via dei Taurini 19, I-00185 Roma, Italy. <sup>3</sup>Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, I-00185 Roma, Italy. Correspondence and requests for materials should be addressed to R.P.-S. (email: romualdo.pastor@upc.edu)

to the properties of the PEV<sup>14</sup>. The localization occurring for large  $\gamma$  implies that the density of infected individuals in the steady state in those processes might not be an extensive quantity, casting doubts on the validity of this theoretical approach and on the actual onset of the endemic infected state.

Here we show that the localization properties of the adjacency matrix PEV for heterogeneous (power-law distributed) networks are described by a picture much more complex than previously believed. In fact, we provide strong numerical evidence that the EC in heterogeneous networks never achieves full delocalization. In the case of uncorrelated synthetic networks with a power-law degree distribution, we obtain, by means of a finite-size scaling analysis, that for mild levels of heterogeneity (with  $\gamma > 5/2$ ), the EC is strongly localized on the hubs, as previously argued. For high heterogeneity ( $\gamma < 5/2$ ), however, we point out that the EC, as measured by the components of the PEV, is highly correlated with the corresponding node's degree. This strong correlation results in an effective localization on a mesoscopic subgraph, that can be identified as the shell with the largest index in the  $K$ -core decomposition of the network<sup>3</sup>. The paper of Goltsev *et al.*<sup>14</sup> is perfectly correct for what concerns the case  $\gamma > 5/2$  but, by only considering the possibility of localization on a finite set of nodes, could not detect the mesoscopic localization occurring for  $\gamma < 5/2$ . In order to overcome the localization effects intrinsic of the EC, a new centrality measure, based on the largest eigenvalue of the Hashimoto, or non-backtracking, matrix, has been recently proposed<sup>16</sup>. We observe that this new centrality is not completely free from localization effects. Thus, while it almost coincides with the EC for  $\gamma < 5/2$ , and for  $\gamma > 5/2$  it avoids the extreme localization around the hubs shown by the EC, it is still localized in this case in some mesoscopic subset of nodes, whose characterization calls for further research. The extension of our analysis to the case of real world networks is hampered by the fact that usually only one network instance is available, which prevents performing a finite-size scaling study. Nevertheless, we numerically argue that also for real networks a twofold scenario holds, in which the PEV is either localized on the hubs, or effectively localized on the maximum  $K$ -core of the network.

## Results

**Eigenvector localization and the inverse participation ratio.** A full characterization of an undirected network of size  $N$  is given by its adjacency matrix<sup>1</sup>  $\mathbf{A}$ , whose elements take the value  $A_{ij} = 1$  if nodes  $i$  and  $j$  are connected by an edge, and value  $A_{ij} = 0$  otherwise. The spectral properties of the adjacency matrix are defined by the set of eigenvalues  $\Lambda_i$ , and associated eigenvectors  $\mathbf{f}(\Lambda_i)$ ,  $i = 1, \dots, N$ , defined by

$$\mathbf{A}\mathbf{f}(\Lambda) = \Lambda\mathbf{f}(\Lambda). \quad (1)$$

Since the adjacency matrix is symmetric all its eigenvalues are real. The largest of those eigenvalues  $\Lambda_1$ , is associated with the principal eigenvector (PEV) which we denote simply by  $\mathbf{f}$ .

The concept of the localization of the PEV  $\mathbf{f}$  translates in determining whether the value of its normalized components is evenly distributed among all nodes in the network, or either it attains a large value on some subset, and is much smaller in all the rest. While this concept is quite easy to grasp, assessing it in a single network instance is a delicate issue because any quantitative definition involves some degree of arbitrariness. The task becomes however straightforward when ensembles of networks of different size can be generated. In such a case, the localization of the eigenvector  $\mathbf{f}$  associated with the eigenvalue  $\Lambda$  can be precisely assessed by computing the inverse participation ratio (IPR), defined as<sup>14,16</sup>,

$$Y_\Lambda = \sum_i f_i^4(\Lambda). \quad (2)$$

In the absence of any knowledge about the localization support, it is possible to determine whether an eigenvector is localized (on some subset in the network) by studying its inverse participation ratio, as a function of the system size  $N$  and fitting its behavior to a power-law decay of the form

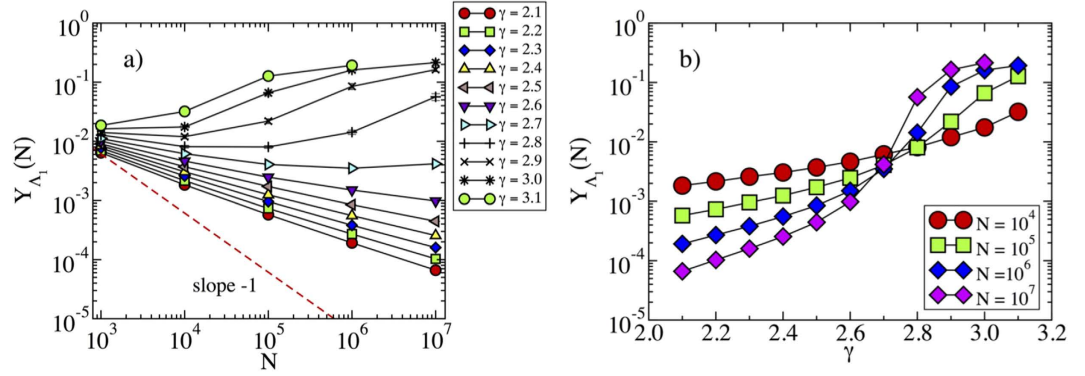
$$Y_\Lambda(N) \sim N^{-\alpha}. \quad (3)$$

If the eigenvector is delocalized, i.e. for  $f_i \sim N^{-1/2}$ , the exponent  $\alpha$  is equal to 1. An exponent  $\alpha < 1$  is evidence that some form of localization is taking place. In the case of extreme localization on a single node, or on a set of nodes with size  $N_V$  independent of the network size  $N$ , the corresponding components of the PEV are finite and this implies  $Y_\Lambda^{\text{loc}} \sim \mathcal{O}(1)$ , i.e.,  $\alpha = 0$  for  $N \rightarrow \infty$ . Finally, if localization takes place over a subextensive set of nodes of size  $N_V \sim N^\beta$ , we expect

$$Y_\Lambda(N) \sim \frac{1}{N_V} \sim N^{-\beta}, \quad (4)$$

leading to a decay exponent  $\alpha = \beta$ .

**Eigenvector localization in synthetic networks.** We study the localization properties of the PEV computed for synthetic power-law distributed networks of growing size, generated using the uncorrelated configuration model (UCM)<sup>18</sup>, a modification of the standard configuration model<sup>19,20</sup> designed to avoid degree correlations<sup>21</sup>. In order to explore the presence or absence of localization, we analyze the scaling of  $Y_\Lambda(N)$  as function of  $N$  as discussed above. In Fig. 1(a) we apply this finite-size scaling analysis to synthetic networks with different values of  $\gamma$ . In this and the following figures, statistical averages are performed over at least 100 different network samples. Error bars are usually smaller than the symbol sizes. In the case of large  $\gamma$  we observe an IPR tending to a constant for large  $N$ , confirming the localization on the hubs predicted by refs. 14,15. The situation is however surprisingly different for  $\gamma < 5/2$ . Thus, while according to Goltsev *et al.*<sup>14</sup>, we should expect a delocalized PEV and an IPR decreasing as  $N^{-\alpha}$  with  $\alpha = 1$ , we observe instead power-law decays with  $N$ , with effective exponents  $\alpha$  always



**Figure 1.** (a) Inverse participation ratio as a function of the network size for the adjacency matrix of synthetic networks with different degree exponent  $\gamma$ . For large  $\gamma$ , the IPR tends to saturate to a constant value for sufficiently large value of  $N$ . For  $\gamma < 5/2$ , on the other hand, the behavior of the IPR can be fitted to  $Y_{\Lambda_1}(N) \sim N^{-\alpha}$ , with  $\alpha < 1/2$ . The dashed line represents a power-law behavior  $\sim N^{-1}$ , corresponding to a delocalized IPR. (b) Inverse participation ratio as a function of the degree exponent  $\gamma$  for different network sizes  $N$ . The plot confirms the presence of transition in the behavior of the IPR, located in the vicinity of  $\gamma = 5/2$ .

smaller than  $1/2$ . The change of behavior of the IPR can be further confirmed in Fig. 1(b), where we plot the IPR as a function of the degree exponent  $\gamma$ , for different values of  $N$ . While it is clear that for  $\gamma \geq 2.7$  the IPR tends to a constant asymptotically, slow crossover effects do not allow to draw firm conclusions based on numerics about the precise value of  $\gamma$  for which the behavior changes. However, since the dependence of the largest eigenvalue on  $N$  changes for  $\gamma = 5/2$ <sup>11</sup> we expect the transition to take place exactly at  $\gamma = 5/2$ : simulation results are perfectly compatible with this result.

The behavior at  $\gamma < 5/2$  can be understood mathematically by observing that the largest eigenvalue in this regime,  $\Lambda_1 = \langle q^2 \rangle / \langle q \rangle$ <sup>11</sup>, coincides with the largest eigenvalue of the adjacency matrix in the annealed network approximation. The annealed network approximation<sup>22,23</sup> consists in replacing the actual, fixed, adjacency matrix by an average performed over degree classes, taking the form

$$\bar{a}_{ij} = \frac{q_j' P(q_i | q_j')}{NP(q_i)}, \tag{5}$$

where  $P(q|q')$  is the conditional probability that a link from a node of degree  $q'$  points to a node of degree  $q$ <sup>24</sup>. For degree uncorrelated networks, with  $P(q|q') = qP(q)/\langle q \rangle$ <sup>25</sup>, we obtain an averaged adjacency matrix

$$\bar{a}_{ij} = \frac{q_i q_j}{N \langle q \rangle}. \tag{6}$$

The matrix  $\bar{a}_{ij}$  is semi-positive definite and therefore all its eigenvalues are non-negative<sup>26</sup>. Then considering that  $\text{Tr}(\bar{\mathbf{a}}^2) = [\text{Tr}(\bar{\mathbf{a}})]^2 = (\langle q^2 \rangle / \langle q \rangle)^2$ , where  $\text{Tr}(\cdot)$  is the trace operator, we have that  $\bar{a}_{ij}$  has a unique non-zero eigenvalue  $\Lambda_{\text{an}} = \langle q^2 \rangle / \langle q \rangle$ , with associated principal eigenvector  $f_i^{\text{an}} \propto q_i$ . Applying the normalization condition  $\sum_i f_i^2 = 1$ , we obtain the normalized form

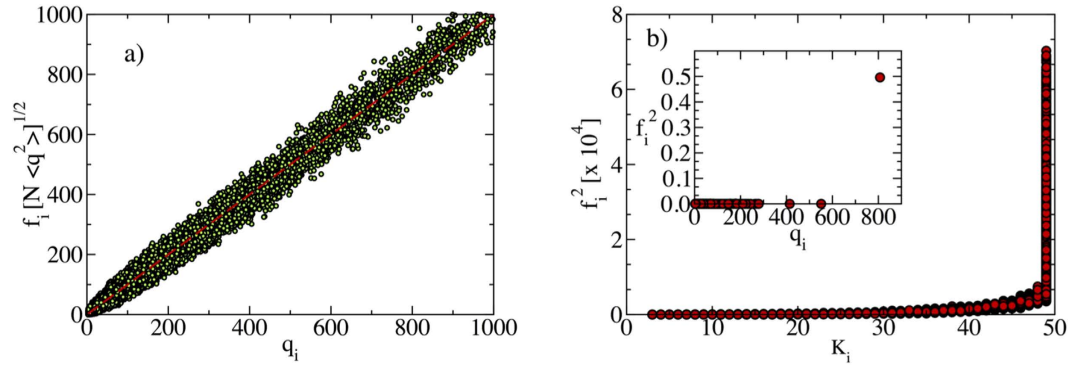
$$f_i^{\text{an}} = \frac{q_i}{[N \langle q^2 \rangle]^{1/2}}. \tag{7}$$

Inserting the expression of  $f_i^{\text{an}}$  into Eq. (2) yields

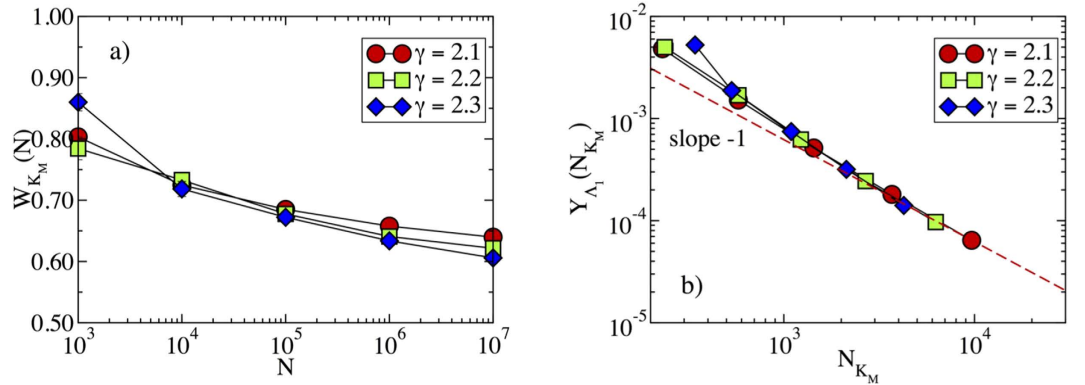
$$Y_{\Lambda_1}(N) \sim 1/N^{(3-\gamma)/2}, \tag{8}$$

that is, a decay with an exponent smaller than  $1/2$ , in agreement with the results in Fig. 1(b). Figure 2(a) confirms that also quenched synthetic networks have PEV components proportional in average to the degree. Notice that Eq. (8) is approximately true only in quenched networks for  $\gamma < 5/2$ , since the condition leading to it, Eq. (7) fails at  $\gamma > 5/2$ , see Fig. 2(b, inset).

A more physical interpretation of the particular distribution of the PEV in power-law networks with  $\gamma < 5/2$ , is that the PEV becomes *effectively* localized on the max(imum)  $K$ -core of the network, defined as the set of nodes with the largest core index  $K_M$  in a  $K$ -core decomposition<sup>3,27</sup>. The  $K$ -core decomposition is an iterative procedure to classify vertices of a network in layers of increasing density of connections. Starting with the full graph, one removes the vertices with degree  $q = 1$ , i.e. with only one connection. This procedure is repeated until only nodes with degree  $q \geq 2$  are left. The removed nodes constitute the  $K = 1$ -shell and those remaining compose the  $K = 2$ -core. At the next step all vertices with degree  $q = 2$  are removed, thus leaving the  $K = 3$ -core. The procedure is repeated iteratively. The maximum  $K$ -core (of index  $K_M$ ) is the set of vertices such that one more iteration of the procedure removes all of them. The line of argument leading to this interpretation stems from combining the



**Figure 2.** (a) Rescaled scatter plot of  $f_i [N \langle q^2 \rangle]^{1/2}$  as a function of  $q_i$  for a synthetic network with  $\gamma = 2.1$  and size  $N = 10^6$ . Data fits the expectation for the PEV in the annealed network approximation, Eq. (7), with only small fluctuations. (b, main) Scatter plot of the squared PEV components as a function of the  $K$ -core index for the adjacency matrix of a power-law synthetic network with  $\gamma = 2.1$  and size  $N = 10^6$ . (b, inset) Scatter plot of the squared PEV components as a function of the degree  $q_i$  in a synthetic network with  $\gamma = 3.5$  and size  $N = 10^6$ .



**Figure 3.** (a) Total weight  $W_{K_M}$  of the PEV on the nodes of the max  $K$ -core in synthetic networks as a function of size  $N$ . (b) Inverse participation ratio as a function of the size of the max  $K$ -core  $N_{K_M}$ . The dashed line represents a power-law behavior  $\sim N_{K_M}^{-1}$ . We can see the asymptotic behavior  $Y_{\Lambda_1} \sim N_{K_M}^{-1}$ , valid for large network sizes.

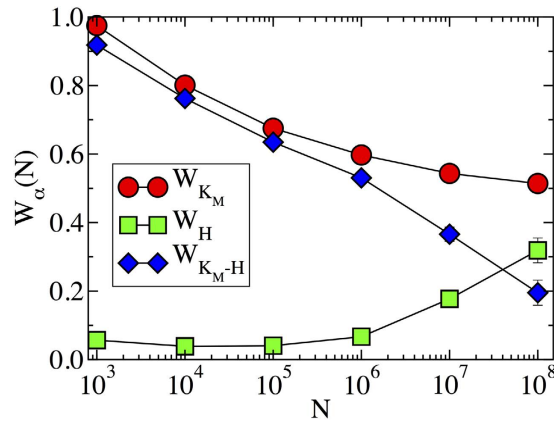
results of ref. 14, in which it is proposed that, in epidemic spreading in complex networks<sup>28</sup>, infection activity is localized on the PEV, with the observations in ref. 29, in which the maximum  $K$ -core is identified as a subset of nodes sustaining epidemic activity for  $\gamma < 5/2$ .

We can see this effective localization on the maximum  $K$ -core in different ways. In the first place, in Fig. 2(b,main) we plot the squared components  $f_i^2$  of the PEV for all vertices against their corresponding  $K$ -core index. From this Figure we conclude that all nodes with the largest  $f_i$  components belong to the max  $K$ -core. The size of this max  $K$ -core,  $N_{K_M}$ , grows sublinearly as a function of the network size as  $N_{K_M} \sim N^{(3-\gamma)/2}$ <sup>27</sup>. However, despite this sublinear growth, a finite fraction of the total PEV weight is concentrated on this subset. We check this fact in Fig. 3(a): the total weight of the nodes in the max  $K$ -core,

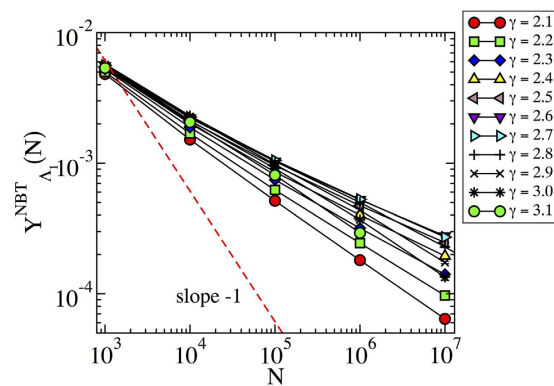
$$W_{K_M} = \sum_{i \in K_M} f_i^2, \tag{9}$$

tends to a constant in the limit of large network size, implying that more than half of the weight of the normalized PEV resides over the max  $K$ -core. Finally, the size dependence of the max  $K$ -core translates, from Eq. (4) in an IPR scaling as  $Y_{\Lambda_1} \sim N_{K_M}^{-1} \sim 1/N^{(3-\gamma)/2}$ , in agreement with the result obtained from the degree dependence of the PEV components,  $f_i \sim q_i$ , see Eq. (8). The relation between IPR and max  $K$ -core size is satisfactorily checked in Fig. 3(b), where we observe it to be valid for large network sizes.

For  $\gamma > 5/2$ , instead, Fig. 2(b,inset) confirms the localization of the PEV around the hub<sup>14,16</sup>, displaying a disproportionately large component on the node with the largest degree. Notice that, irrespective of the value of  $\gamma$ , with high probability the hub belongs to the max  $K$ -core. What changes in the two cases is that for  $\gamma > 5/2$  the hub alone carries a finite fraction of the normalization weight ( $f_i \sim \mathcal{O}(1)$ ) while for  $\gamma < 5/2$  it carries a vanishing fraction, and all nodes of the max  $K$ -core must be considered to have a finite weight  $W_{K_M}$ . The behavior for  $\gamma > 3$  is clearly evident from Fig. 2(b,inset). In the case  $5/2 < \gamma < 3$ , the accumulation of a finite weight on the hub takes place for



**Figure 4.** Weight of the PEV as a function of the network size in power-law networks with degree exponent  $\gamma = 2.8$ . The different functions correspond to: total weight of the nodes in max  $K$ -core,  $W_{K_M}$ ; total weight in the hub,  $W_H$ ; total weight in the max  $K$ -core, subtracting the hub,  $W_{K_M-H}$ .



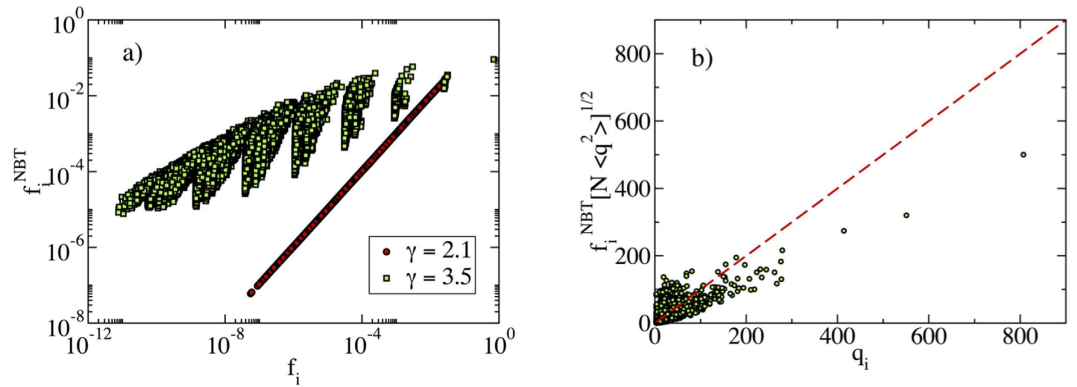
**Figure 5.** Inverse participation ratio as a function of the network size  $N$  for the NBTC for power-law synthetic networks with different degree exponents  $\gamma$ . The dashed line has slope  $-1$  indicating delocalization.

sufficiently large  $N$ . This effect is observed in Fig. 4, where we plot the total weight  $W_{K_M}$  of the nodes in the max  $K$ -core, Eq. (9), the total weight in the hub,  $W_H$ , and the total weight in the max  $K$ -core, subtracting the hub,  $W_{K_M-H}$ . As we can observe from this Figure, the weight at the hub is small for network sizes  $N < 10^6$ , but it then starts to increase, to finally take over, for large network sizes  $N > 10^7$ .

**The non-backtracking centrality.** The observations presented here, together with the arguments provided by Martin *et al.*<sup>16</sup>, hint that the EC is problematic as a useful measure of centrality. For large values of  $\gamma$ , it is affected by an exceedingly strong localization on the hub, arising as a purely topological artifact: the hub is central because its neighbors are central, but those in turn are central only because of the hub. For small values of  $\gamma$ , on the other hand, the observed relation  $f_i \sim q_i$  indicates that the eigenvector centrality provides essentially the same information as the degree centrality. As an attempt to correct the flaws of the EC, Martin *et al.*<sup>16</sup> propose a modified centrality measure, the non-backtracking centrality (NBTC), which is computed in terms of the non-backtracking matrix. The Hashimoto, or non-backtracking matrix (NBT)<sup>16,30,31</sup>, is defined as follows: an initially undirected network is converted into a directed one by transforming each undirected edge into a pair of directed edges, each pointing in opposite directions. If the initial undirected network has  $E$  edges, the NBT matrix is a  $2E \times 2E$  matrix with rows and columns corresponding to directed edges  $i \rightarrow j$  with value  $B_{i \rightarrow j, l \rightarrow m} = \delta_{i,m}(1 - \delta_{j,l})$ ,  $\delta_{i,j}$  being the Kronecker symbol. The components of the principal eigenvector of the NBT matrix,  $f_{i \rightarrow j}$  measure the centrality of vertex  $i$  disregarding the contribution of vertex  $j$ . The NBT centrality of vertex  $j$  is given by the sum of these contributions for all neighbors of  $j$ :  $f_j^{NBT} = \sum_i A_{ij} f_{i \rightarrow j}$ . The elements of the NBT matrix count the number of non-backtracking walks in a graph and hence remove self-feedback in the calculation of node centrality, thus eliminating in principle the artificial topological enhancement of the hub's centrality.

As Fig. 5 shows, however, the NBTC is not free from localization effects: For all values of  $\gamma$  the NBTC is not delocalized, i.e.  $Y_{\Lambda_1}$  does not decrease as  $1/N$  when increasing  $N$ . This fact can be understood for  $\gamma < 5/2$  in view of the previous results. The adjacency matrix PEV is localized on the max  $K$ -core, which features many mutual interconnections: the centrality of a node is only weakly affected by self-feedback, and removing the contribution of backtracking paths has therefore little effect. This is confirmed by the scatter plot of the NBTC values  $f_i^{NBT}$  as a function of the corresponding components  $f_i$  of the adjacency matrix PEV, computed for the same synthetic





**Figure 6.** (a) Scatter plot of the NBTC centralities  $f_i^{\text{NBT}}$  as a function of the corresponding components of the PEV of the adjacency matrix  $f_i$ , in synthetic uncorrelated networks with a power-law degree distribution. Network size  $N = 10^6$ . (b) Rescaled scatter plot of the NBTC centralities  $f_i^{\text{NBT}} [N \langle q^2 \rangle]^{1/2}$  as a function of  $q_i$  for a synthetic network with  $\gamma = 3.5$  and size  $N = 10^6$ .

networks, Fig. 6(a). For  $\gamma < 5/2$  the two quantities are very strongly correlated. For  $\gamma > 5/2$  instead, Fig. 6(a) shows that the NBT centrality is truly different and uncorrelated from the adjacency matrix EC. However, as Fig. 5 shows, the NBT IPR, computed from the components of the NBTC, decreases with the system size  $N$  more slowly than  $N^{-1}$ . This is indicative that also in this case a localization occurs on a mesoscopic subset, whose size grows sublinearly. Figure 6(b) shows that this localization is not due to a strong correlation between the NBT centrality and the degree of nodes, contrary to what happens for the EC for  $\gamma < 5/2$ .

**Eigenvector localization in real networks.** For real networks, which have fixed size and do not allow for a finite size scaling analysis, localization is necessarily a more blurred concept. The value of  $Y_{\Lambda_1}$  gauges how localized the PEV is, but it does not permit to unambiguously declare a network localized or not. However, also in this case it is possible to detect, as in synthetic networks, the existence of different localization modes. We consider here several real complex networks exhibiting large variations in size, heterogeneity and degree correlations (see Methods and Supplemental Material, SM, for details).

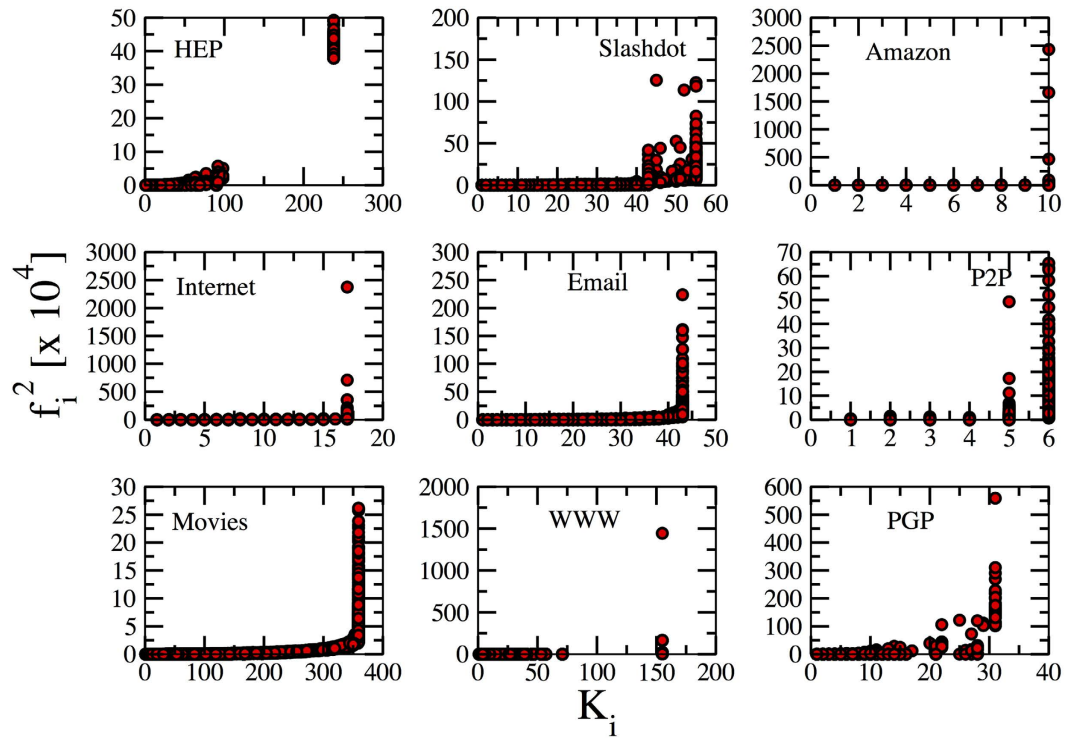
The linear relation between  $f_i$  and the degree  $q_i$  is not fulfilled in real networks (see Supplementary Figure SF-1), probably due to the presence of nontrivial degree correlations (see SM) which are absent in the synthetic networks. The effective localization on the max- $K$  core is however still present in some cases. In Fig. 7 we plot for these networks the squared PEV component  $f_i^2$  as a function of the  $K$ -core index. In some cases (HEP, Movies) all nodes in the max  $K$ -core have a comparable and large EC (as in synthetic networks for  $\gamma < 5/2$ ), suggesting localization on the max  $K$ -core. In other cases (Internet, Amazon) one or a few nodes have a disproportionately large value of  $f_i^2$ , hinting at a localization around hubs, as in synthetic networks for large  $\gamma$ .

To clarify the phenomenology we report in Table 1 for each of the real-world networks the values of the leading eigenvalue, and the factors  $\langle q^2 \rangle / \langle q \rangle$  and  $\sqrt{q_{\max}}$ . The analysis here is complicated by the presence of degree correlations (see SM), which invalidate the direct connection<sup>11</sup> between  $\Lambda_1$  and the largest between  $\sqrt{q_{\max}}$  and  $\langle q^2 \rangle / \langle q \rangle$ <sup>14</sup>. However, in some cases (Internet, Amazon) the leading eigenvalue is much closer to  $\sqrt{q_{\max}}$  than to  $\langle q^2 \rangle / \langle q \rangle$ : This suggests a localization around the hub and matches well with Fig. 7. In others the opposite is true:  $\Lambda_1$  is very far from  $\sqrt{q_{\max}}$  and relatively close to  $\langle q^2 \rangle / \langle q \rangle$ , hinting at a localization on the max  $K$ -core, again in agreement with Fig. 7. In other cases (P2P, WWW), values are so close that no conclusion can be drawn.

A further confirmation of this picture is provided by the analysis of the NBT centrality. When localization occurs on hubs one expects the elimination of backtracking paths to have a strong impact, as self-feedback effects are tamed. In this case we expect the ratio between the IPR for the NBTC and the IPR for the adjacency matrix to be small. On the contrary, when the localization occurs on the max  $K$ -core, passing from the adjacency to the NBT matrix would not lead to a big change and we expect the ratio to be close to 1. Table 1 confirms this expectation: the IPR ratio is small when the leading eigenvalue  $\Lambda_1$  is essentially given by  $\sqrt{q_{\max}}$  (localization on hubs) while it is close to 1 when  $\Lambda_1$  is closer to the  $\langle q^2 \rangle / \langle q \rangle$  factor (localization on the max  $K$ -core). A visual representation of these results is provided in Fig. 8, where we plot the IPR ratio as a function of the ratio between  $\Lambda_1$  and  $\sqrt{q_{\max}}$ . As we can see, networks in which the PEV is localized in the max  $K$ -core are situated in the upper right corner of the panel, while the lower left corner shows the networks with localization occurring on the hubs.

## Discussion

The properties of the principal eigenvector (PEV), and associated largest eigenvalue, of the adjacency matrix defining a network have a notable relevance as characterizing several features of its structure and its effects on the behavior of dynamical processes running on top of it. Most important among these features is the role of the components of the PEV as a measure of a node's importance, the so-called eigenvector centrality. One of the properties of the PEV that has recently attracted the interest of the statistical physics community is its localization. In the case of networks with a power-law degree distribution  $P(q) \sim q^{-\gamma}$ , initial research on this subject<sup>14,16</sup> suggested



**Figure 7.** Scatter plot of squared PEV components of the adjacency matrix of the real-world networks as a function of the  $K$ -core index.

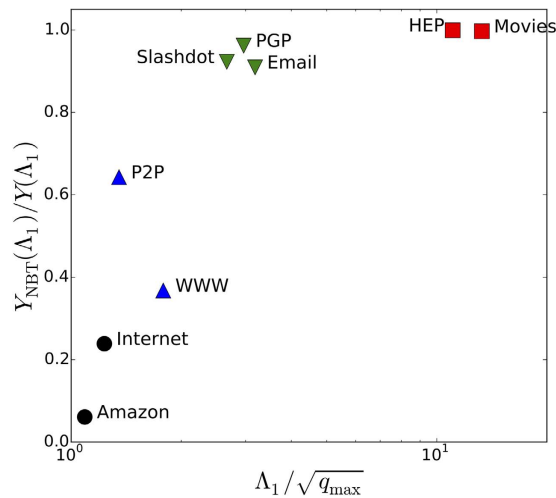
Network	$N$	$\langle q \rangle$	$\langle q^2 \rangle / \langle q \rangle$	$\sqrt{q_{\max}}$	$\Lambda_1$	$Y(\Lambda_1)$	$Y(\Lambda_1^{NBT})$	IPR Ratio
HEP	12006	19.74	129.94	22.16	244.93	0.003890	0.003887	0.9993
Slashdot	82168	12.27	149.71	50.52	134.63	0.002174	0.002006	0.9228
Amazon	403394	12.11	30.55	52.46	57.15	0.089122	0.005423	0.0608
Internet	10790	4.16	259.46	48.34	59.58	0.066138	0.015783	0.2386
Email	36692	10.02	140.08	37.19	118.42	0.003790	0.003446	0.9091
P2P	62586	4.73	11.60	9.75	13.18	0.000921	0.000592	0.6429
Movies	81860	89.53	594.92	61.55	817.36	0.000640	0.000638	0.9966
WWW	325729	6.69	280.68	103.54	184.93	0.022726	0.008357	0.3677
PGP	10680	4.55	18.88	14.32	42.44	0.016622	0.015989	0.9619

**Table 1.** Relevant metrics for the various real-world networks with and the measured value of the IPR ratio between  $Y(\Lambda_1^{NBT})$  and  $Y(\Lambda_1)$ . Size and other information on the networks are provided in the Supplementary Information.

that, for  $\gamma > 5/2$ , the PEV is localized on the nodes with largest degree. On the other hand, for  $\gamma < 5/2$ , the PEV should be delocalized.

In this paper we have shown that eigenvector localization in heterogeneous networks is described by a more complex picture. Thus, we present evidence that for all power-law distributed networks the PEV is always localized to some extent. In the case of synthetic power-law distributed networks, we observe that, while for mildly heterogeneous networks with  $\gamma > 5/2$  the PEV is indeed localized on the nodes with maximum degree (the hubs), in the case of high heterogeneity, with  $\gamma < 5/2$ , the PEV shows a peculiar form of localization, its components  $f_i$  being proportional to the node's degree,  $f_i \sim q_i$ . This particular proportionality induces an effective localization on the maximum  $K$ -core of the network, defined as the core of maximum index in a  $K$ -core decomposition. This max  $K$ -core concentrates a finite fraction of the normalized weight of the PEV, despite the fact that the size of the max  $K$ -core is sublinear with the network size. In the case of real world networks, the elucidation of the PEV localization is not so clearcut. We however provide evidence for an analogous scenario as that observed in synthetic networks, where the nature of the localization of the PEV is ruled by its associated largest eigenvalue  $\Lambda_1$ : When  $\Lambda_1$  is close to the mean-field value  $\langle q^2 \rangle / \langle q \rangle$ , localization on the max  $K$ -core is expected. On the other hand, when the largest eigenvalue is close to  $\sqrt{q_{\max}}$ , localization takes place on the hubs.

The results presented here give a new perspective on complex topologies from several viewpoints. Firstly, it is common knowledge that networks with  $\gamma > 3$  are fundamentally different from those with  $\gamma < 3$  (scale-free networks) because the divergence of the second moment of the degree distribution has a series of crucial effects. A



**Figure 8. Ratio between the NBTC IPR and the IPR of the adjacency matrix as a function of the ratio between the largest eigenvalue and the square root of the maximum degree, for the real networks considered.** The symbol codes are: square for localization on the max  $K$ -core; circle for localization on the hub; triangle up for networks in which  $\sqrt{q_{\max}}$  is very close to  $\langle q^2 \rangle / \langle q \rangle$ , so no conclusion can be drawn; triangle down for the rest of networks.

tacit corollary is that networks with  $2 < \gamma < 3$  have essentially the same properties. Our paper, together with other recent results<sup>14</sup>, points out that networks with exponent  $\gamma < 5/2$  are in many respects qualitatively different from those with  $\gamma > 5/2$ . Secondly, our results point out the weakness of eigenvector centrality as a measure of centrality for power-law networks. Indeed, for  $\gamma < 5/2$ , eigenvector centrality does not provide more information than degree centrality, while for  $\gamma > 5/2$  the eigenvector localization on the hubs arises as a purely topological artifact. Alternative measures of centrality, based on the Hashimoto non-backtracking matrix<sup>16,30,31</sup> are also not free from localization effects. Finally, from a dynamical point of view, largest eigenvalues and the associated eigenvectors are crucially related to the properties of processes on networks<sup>7,14,32</sup> and their localization effects should be taken properly into account when developing theories relying on the structure of the adjacency matrix.

The localization properties described here call for a revision of our present understanding of heterogeneous topologies. Other networks properties, such as degree correlations, clustering or the presence of a community structure, might play a role in the localization of the PEV. The clarification of these effects, as well as the understanding of the nature of the mesoscopic subgraph on which the NBTC is localized for  $\gamma > 5/2$ , are still open questions, calling for further scientific effort.

## Methods

**Real networks analyzed.** We consider in our analysis the following real networks datasets:

- **HEP:** Collaboration network between authors of papers submitted to the High Energy Physics section of the online preprint server arXiv. Each node is a scientist. Two scientists are connected by an edge if they have coauthored a preprint<sup>33</sup>.
- **Slashdot:** User network of the Slashdot technology news website. Nodes represent users, which can tag each other as friends or foes. An edge represents the presence of a tagging between two users<sup>34</sup>.
- **Amazon:** Co-purchasing network from the online store Amazon. Nodes represent products, which are joined by edges if they are frequently purchased together<sup>35</sup>.
- **Internet:** Internet map at the Autonomous System level, collected at the Oregon route server. Vertices represent autonomous systems (aggregations of Internet routers under the same administrative policy), while edges represent the existence of border gateway protocol (BGP) peer connections between the corresponding autonomous systems<sup>36</sup>.
- **Email:** Enron email communication network. Nodes represent email addresses. An edge joins two addresses if they have exchanged at least one email<sup>34</sup>.
- **P2P:** Gnutella peer-to-peer file sharing network. Nodes represent hosts in the Gnutella system. An edge stands for a connection between two Gnutella hosts<sup>33</sup>.
- **Movies:** Network of movie actor collaborations obtained from the Internet Movie Database (IMDB). Each vertex represents an actor. Two actors are joined by an edge if they have co-starred at least one movie<sup>37</sup>.
- **WWW:** Notre Dame web graph. Nodes represent web pages from University of Notre Dame. Edges indicate the presence of a hyperlink pointing from one page to another<sup>38</sup>.
- **PGP:** Social network defined by the users of the pretty-good-privacy (PGP) encryption algorithm for secure information exchange. Vertices represent users of the PGP algorithm. An edge between two vertices indicates that each user has signed the encryption key of the other<sup>39</sup>.

Some of this networks are actually directed. We have symmetrized them, rendering them undirected, to perform our analyses.



## References

- Newman, M. E. J. *Networks: An introduction* (Oxford University Press, Oxford, 2010).
- Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1977).
- Siedman, S. B. Network structure and minimum degree. *Social Networks* **5**, 269–287 (1983).
- Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology* **2**, 113–120 (1972).
- Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43 (1953).
- Brin, S. & Page, L. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* **30**, 107–117 (1998).
- Restrepo, J. G., Ott, E. & Hunt, B. R. Onset of synchronization in large networks of coupled oscillators. *Phys. Rev. E* **71**, 036151 (2005).
- Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J. & Faloutsos, C. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.* **10**, 1–26 (2008).
- Castellano, C. & Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701 (2010).
- Farkas, I. J., Derényi, I., Barabási, A.-L. & Vicsek, T. Spectra of “real-world” graphs: Beyond the semicircle law. *Phys. Rev. E* **64**, 026704 (2001).
- Chung, F., Lu, L. & Vu, V. Spectra of random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **100**, 6313–6318 (2003).
- Dorogovtsev, S. N., Goltsev, A. V., Mendes, J. F. F. & Samukhin, A. N. Spectra of complex networks. *Phys. Rev. E* **68**, 046109 (2003).
- Kühn, R. Spectra of sparse random matrices. *Journal of Physics A: Mathematical and Theoretical* **41**, 295002 (2008).
- Goltsev, A. V., Dorogovtsev, S. N., Oliveira, J. G. & Mendes, J. F. F. Localization and Spreading of Diseases in Complex Networks. *Phys. Rev. Lett.* **109**, 128702 (2012).
- Nadakuditi, R. R. & Newman, M. E. J. Spectra of random graphs with arbitrary expected degrees. *Phys. Rev. E* **87**, 012803 (2013).
- Martin, T., Zhang, X. & Newman, M. E. J. Localization and centrality in networks. *Phys. Rev. E* **90**, 052808 (2014).
- Ferreira, S. C., Castellano, C. & Pastor-Satorras, R. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Phys. Rev. E* **86**, 041125 (2012).
- Catanzaro, M., Boguñá, M. & Pastor-Satorras, R. Generation of uncorrelated random scale-free networks. *Phys. Rev. E* **71**, 027103 (2005).
- Molloy, M. & Reed, B. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* **6**, 161 (1995).
- Molloy, M. & Reed, B. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probab. Comput.* **7**, 295 (1998).
- Boguñá, M., Pastor-Satorras, R. & Vespignani, A. Cut-offs and finite size effects in scale-free networks. *Euro. Phys. J. B* **38**, 205–210 (2004).
- Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. Critical phenomena in complex networks. *Rev. Mod. Phys.* **80**, 1275–1335 (2008).
- Boguñá, M., Castellano, C. & Pastor-Satorras, R. Langevin approach for the dynamics of the contact process on annealed scale-free networks. *Phys. Rev. E* **79**, 036110 (2009).
- Pastor-Satorras, R., Vázquez, A. & Vespignani, A. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* **87**, 258701 (2001).
- Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks. *Advances in Physics* **51**, 1079–1187 (2002).
- Gantmacher, F. R. *The theory of matrices*, vol. II (Chelsea Publishing Company, New York, 1974).
- Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. *k*-core organization of complex networks. *Phys. Rev. Lett.* **96**, 040601 (2006).
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
- Castellano, C. & Pastor-Satorras, R. Competing activation mechanisms in epidemics on networks. *Nature Scientific Reports* **2**, 371 (2012).
- Hashimoto, K. Zeta functions of finite graphs and representations of *p*-adic groups. *Adv. Stud. Pure Math.* **15**, 211–280 (1989).
- Krzakala, F. et al. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences* **110**, 20935–20940 (2013).
- Karrer, B., Newman, M. E. J. & Zdeborová, L. Percolation on sparse networks. *Phys. Rev. Lett.* **113**, 208702 (2014).
- Kleinberg, J. L. J. & Faloutsos, C. Graph evolution: Densification and shrinking diameter. *ACM Trans. Knowl. Discov. Data* **1**, 2 (2007).
- Leskovec, J., Lang, K., Dasgupta, A. & Mahoney, M. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* **6**, 29–123 (2009).
- Leskovec, J., Adamic, L. A. & Huberman, B. A. The dynamics of viral marketing. *ACM Trans. Web* **1** (2007).
- Pastor-Satorras, R. & Vespignani, A. *Evolution and structure of the Internet: A statistical physics approach* (Cambridge University Press, Cambridge, 2004).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Albert, R., Jeong, H. & Barabási, A.-L. Diameter of the World-Wide Web. *Nature* **401**, 130–131 (1999).
- Boguñá, M., Pastor-Satorras, R., Daz-Guilera, A. & Arenas, A. Models of social networks based on social distance attachment. *Phys. Rev. E* **70**, 056122 (2004).

## Acknowledgements

R.P.-S. acknowledges financial support from the Spanish MINECO, under projects No. FIS2010-21781-C02-01 and FIS2013-47282-C2-2, EC FET-Proactive Project MULTIPLEX (Grant No. 317532), and ICREA Academia, funded by the Generalitat de Catalunya.

## Author Contributions

R.P.-S. and C.C. designed the research. R.P.-S. performed the data analysis. R.P.-S. and C.C. wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Pastor-Satorras, R. and Castellano, C. Distinct types of eigenvector localization in networks. *Sci. Rep.* **6**, 18847; doi: 10.1038/srep18847 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>