# An improved eigenvector centrality based on the non-backtracking matrix

Travis Martin,[1] Xiao Zhang,[2] and M. E. J. Newman[3]
[1]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109
[2]Department of Physics, University of Michigan, Ann Arbor, MI 48109
[3]Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109

We show that eigenvector centrality behaves poorly on sparse random networks with vertices of unusually high degree, or hubs. We give numerical and analytical evidence that traditional eigenvector centrality is susceptible to localization, giving disproportionately high centrality to hubs and their neighbors. We propose an alternative centrality measure, *non-backtracking centrality*, which converges to eigenvector centrality for dense networks and avoids the problem of localization around hubs in sparse networks.

Eigenvector centrality is popular and has many uses.

It is based on the leading eigenvector, which has been studied in depth. Chung *et al.* [1] lower bounded it by the weighted average of the squares of node degrees and by the square root of the largest degree. Many sparse graphs have little degree variation, and thus their leading eigenvector is a global measure corresponding to the weighted average of squared degrees, and is robust to local changes.

However, the leading eigenvalue is also bounded by the square root of the largest degree. For networks with unbalanced enough degree distributions, the leading eigenvalue corresponds to the largest hub in the network. Nadakuditi and Newman [2] analyze this phenomenon for Poisson random graphs with a planted hub. They show the existance of an eigenvalue for the hub vertex, and show that as the hub grows, its eigenvalue overtakes the average-degree eigenvalue.

In Figure 1 we show how dramatic this can be. We show the same Poisson random graph with an added hub of size 40, 80, and 120, where the area of each node is proportional to its eigenvector centrality on the left. By the third panel the eigenvalue of the hub has become the leading eigenvalue, and the hub's eigenvector centrality is much larger. The right hand side illustrates our proposed solution to this localization, non-backtracking centrality. It shows the exact same graphs, but with node size proportional to non-backtracking centrality. We see more reasonable hub behavoior in this case.

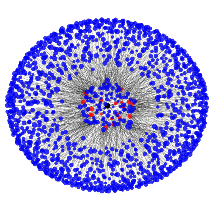

FIG. 1: Transition for the same Poisson random graph with growing hub size.

In this paper we follow Nadakuditi and Newman [2] and let our network be a Poisson random graph of expected degree $c$, plus a single additional hub of degree $k_n$. There is an eigenvalue corresponding to the average degree, $c+1$ and an eigenvalue corresponding to the hub, $k_n/\sqrt{k_n - c}$. We find that these cross over, and thus find a transition in the behavior of the eigenvector, when

$$\frac{k_n}{\sqrt{k_n - c}} = c + 1, \tag{1}$$

or $k_n = c(c+1)$.

After the cross over, there is localization around the hub eigenvector. The hub element is $O(1)$ and its neighbors are $O(...)$, and all other nodes are $O(...)$. We compute an order parameter to measure the degree of localization, which is:

$$\omega = \sqrt[4]{\sum_i^n v_i^4} \tag{2}$$

$\omega$ was chosen so that the most localized eigenvector, a vector with one element equal to 1 and the rest equal to 0, gives $\omega = 1$ and the least localized eigenvector, a vector with each element $= 1/\sqrt{n}$, gives $\omega = 1/\sqrt[4]{n}$, which tends to 0 in the limit of large network size.

Simulation results match nearly identically to the theoretical results. In Figure 2 we see a clear crossing of eigenvalues at $k_n = 110$, as predicted. We also see a sharp phase transition in the order parameter at this point in Figure 3, as the leading eigenvector elements transition from being roughly uniform to dominated by the single element corresponding to the hub.

We also find similar results in more realistic network models. Chung *et al.* [1] analyze the leading eigenvalue of networks generated by the configuration model with power law degree distributions of different exponents, and demonstrate a transition from an eigenvalue corresponding to the average squared degree to an eigenvalue corresponding to the square root of the maximum degree, at an exponent of 2.5. In Figure 4 we see this corresponds to an increase in order parameter, despite the fact that a higher power law exponent is representative of a less uniform degree distribution. This suggests that real-world networks with power-law exponent greater than 2.5 can expect localization around the largest network hub. This might affect algorithms which depend on eigenector-centrality related measures, such as PageRank.

The underlying cause of eigenvector centrality's undesirable properties is an echo chamber effect resulting from each node's centrality being influenced by its neighbors centralities. Each of the hub's neighbors obtain high centrality because they are connected to the hub, which has high centrality. The hub obtains high centrality because it has high degree, and also because it is connected to many nodes which are deemed central on account of their being connected to the hub. This amplification of centrality is what allows for extreme localization, and can be avoided by a centrality measure which takes into account neighbor centrality, but only neighbor centrality derived from nodes *other than oneself.*

This concept is perfectly encapsulated by the non-backtracking matrix, or Hashimoto matrix, which has been employed and analyzed in a variety of contexts [3**?** , 4]. The non-backtracking matrix is a $2m \times 2m$ representation of a graph in which undirected graph is turned into two directed edges, and edge $(i, j)$ is connected to every edge leaving $j$ except for its reciprocal, $(j, i)$.

This forbidding of backtracking prevents the echo chamber effect of eigenvector centrality, and solves all of the problems we've seen so far. (We analyze each of the figures and show how much more consistent it is).

We can see that non-backtracking centrality largely corresponds to eigenvector centrality. This correspondance is exact in the limit of large networks, as the one excluded edge from each edge is asymptotically negligible.

A $2m \times 2m$ matrix can be significantly larger than the original $n \times n$ adjacency matrix, and could raise concerns of intractibility. As shown in by Krzakala *et al.* [4], it is possible to compute the non-backtracking centrality of each node using only a $2n \times 2n$ matrix (show the matrix).

This centrality measure is great, but certainly not an end-all be-all. For example, it doesn't work on trees. The problem of reasonable behavior on trees has yet to be solved. We also believe there are related problems here, including belief propagation and the computation of core-periphery structure.

[1–4]

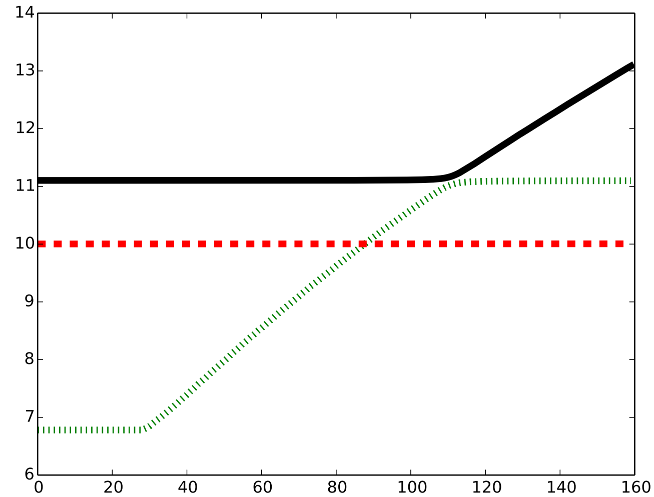


FIG. 2: Eigenvalues for a graph with growing hub.

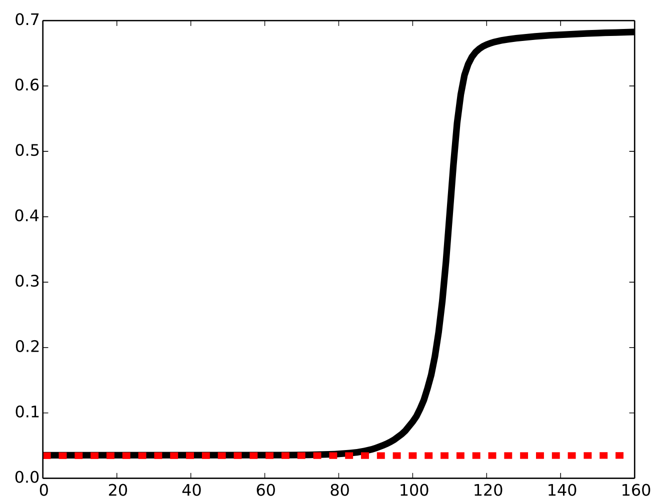


FIG. 3: Order parameter a graph with growing hub.

[1] F. Chung, L. Lu, and V. Vu, Proceedings of the National Academy of Sciences **100**, 6313 (2003).

[2] R. R. Nadakuditi and M. E. Newman, Physical Review E **87**, 012803 (2013).

[3] K.-i. Hashimoto, Automorphic forms and geometry of arithmetic varieties. pp. 211–280 (1989).

[4] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, arXiv preprint arXiv:1306.5550 (2013).

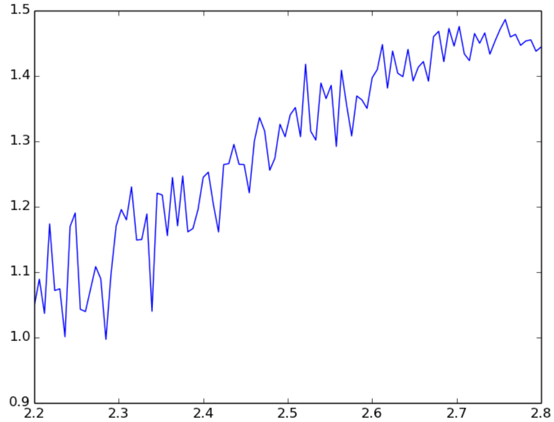


FIG. 4: Order parameter for a power law graph.