Commentaries on Albert-László Barabási's books

Networks101Link5.1   Hubs ©frederick david abraham, 24 Feb 2013

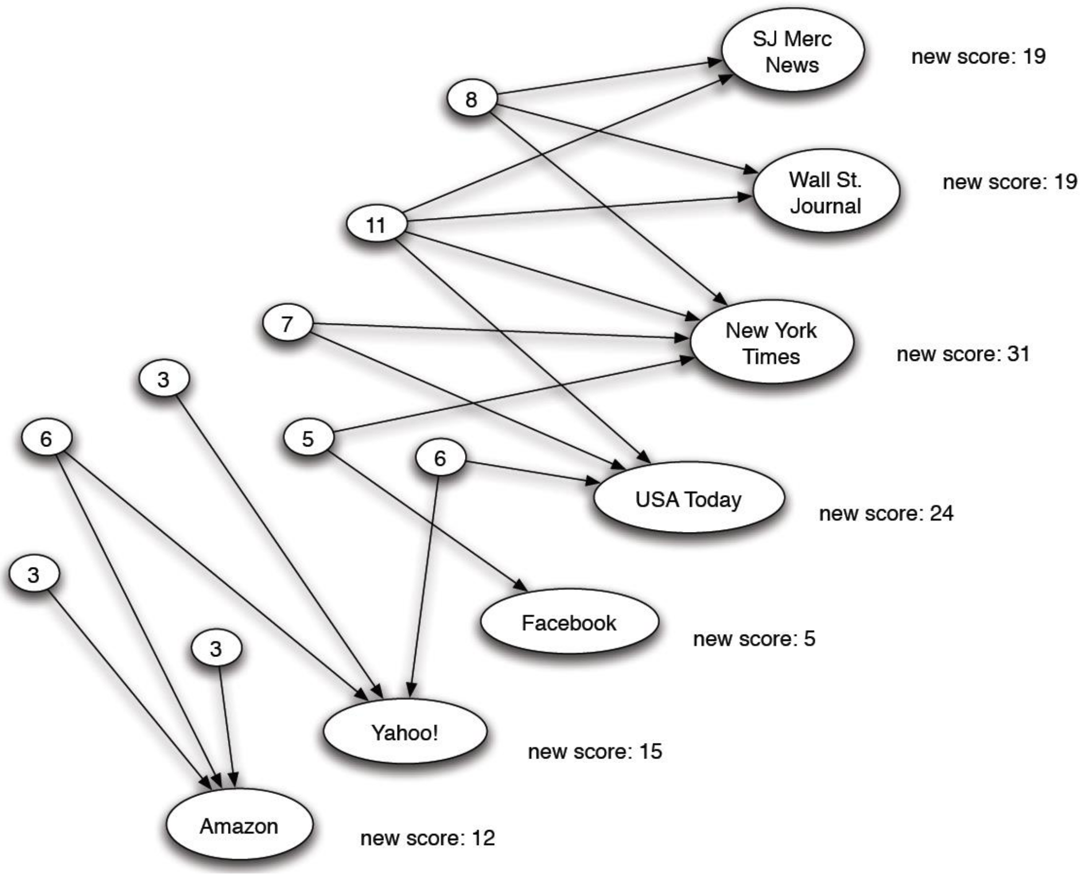
At first I thought there was little to mull over regarding Barabási's simple narrative. Instead of a weak connector between clusters as in small worlds, hubs are a super connector between many clusters. Thus there is an order of magnitude of the shortening of communication among remote nodes of distributed in many clusters. Is this a bifurcation to a new class of models. Obviously there is a continuum of degree of such connectors, so how does one know if this is a real bifurcation and if so, on what parameters does it depend? These questions kind of take us ahead of the story because he has just introduced the concept, and one expects the rest of the book is to be an elaborative trajectory from here.

There is a major difference between the more closed-like system of Cellular Automata (CA) and those of the more open nature of the more open system like scale free and hub dominated ones in terms of their inclusion of changes in nature and number of nodes and links.

At first I got sidetracked into thinking the main parameter might be degree, but it should have been obvious that could not be it, for one could have a node of high degree about which a cluster is huddled. But I did investigate it just a bit looking at three of my model data-sets, elegans[i], Val, and LesMiserables. One thing that Gephi provides, besides running a large number of network measures and images, are tables of node properties, so having thought degree was important, I sorted nodes on the basis of in-degree, out-degree, and total degree. The results were quite interesting despite their irrelevance to the hubbiness issue.

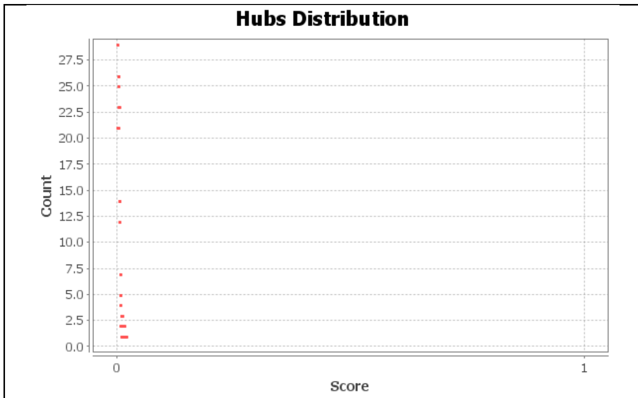
At first I thought in-degree might be most important (in CA's such as Boolean KN Networks), K stands for in degree, suggesting their dominant role in network behavior) and interestingly comparison of in-degree ranking and out-degree ranking varied in their choices of the most important (largest). So I looked at total degree, and with the total equaling the sum of both types of degree, the largest included those of high in-degree, some of high out-degree, and some of nearly equal ins and outs. I found pretty much the same result for a network of jazz musicians.[ii] If one starts to imagine social networks, one could readily imagine those in which inputs are more important, outputs as more important, and two way interaction as being most important. It all depends. There are also a host of other parameters of links that could obviously vary. So I gave up trying to reinvent the wheel or its spokes, and instead of waiting for the next chapters, I grabbed the text-book living in my computer for definitions of Hubs. I was buried in chapter 14, 400 pages in, so you know I didn't have a prayer of understanding it.

So here is my take on Kleinberg's method of ranking hubs as best I can understand it.[iii] Supposing there exist a set of hubs with directed links coming in from another set of hubs. You could rank the receiving set by in-degree. Can you improve on that ranking depending on the value of the hubs whose links supply those hubs? Kleinberg does this by developing two masures, a value measure and a hub measure, illustrated by this figure of.

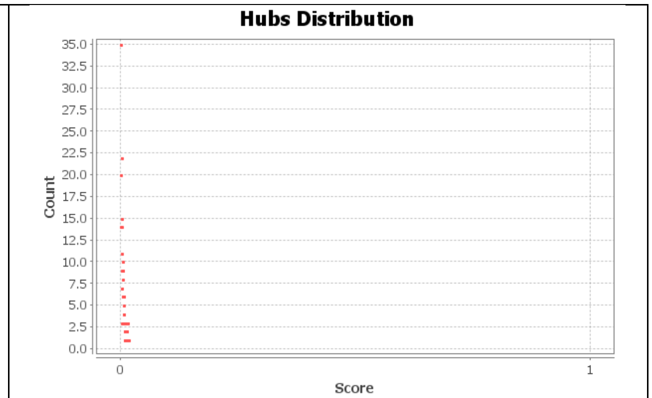


If you took the in-degree to rank the importance of the nodes on the right, you get a very different picture than the 'new scores shown here. The nodes on the left have an authority score that is the sum of all the in-degrees of the nodes to which it is linked, reflecting the idea that they are more important to the extent that they send to more nodes. These value scores are added up for a new score for the receiving nodes as the sum of all the value scores of the nodes from which their in-links come. Thus two scores for each node. Gephi uses Kleinberg's algorithm in its measure 'Hits' which gives both scores for all nodes.[iv]

First, the rankings based on degree were pretty similar to those I mentioned before for elegans, Val, and jazz musicians, the top scores were dominated by in-degree in some cases, out-degree in others, and more equally in some. However, ranking by either the authority (value) or hub measure (which were highly correlated with each other) were dominated by the in-degree count. Browsing the usual places on the internet for 'hubs' shows many sites designating degree, and some, in-degree as important, the latter being supported by this result, which means that the Kleinberg algorithm, while more nuanced by using two tiers of in-degree, weights the in-degree most heavily. Here are the distribution plots for the hub scores reported by Gephi for elegans and Val, showing their great similarity, and while I haven't run them as log functions, one would suspect that they are both scale-free with similar slopes.



Caenhabtitus elegans



Val FaceBook

End Notes

---

[i] C. Elegans neural network: A directed, weighted network representing the neural network of C. Elegans. Data compiled by D. Watts and S. Strogatz and made available on the web here. Please cite D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998). Original experimental data taken from J. G. White, E. Southgate, J. N. Thompson, and S. Brenner, Phil. Trans. R. Soc. London 314, 1-340 (1986).

[ii] List of edges of the network of Jazz musicians. P.Gleiser and L. Danon , Adv. Complex Syst.6, 565 (2003).

[iii] Easley, D., & Kleinberg, J (2010). *Networks, Crowds, & Markets: Reasoning about a Highly Connected World*. Pp. 399 inc. Fig. 14.3.

[iv] Jon M. Kleinberg, Authoritative Sources in a Hyperlinked Environment, in *Journal of the ACM 46 (5):* 604–632 (1999)