On the Eigenvalue Power Law

Milena Mihail¹
Christos Papadimitriou²

¹ College of Computing,
Georgia Institute of Technology
Atlanta, GA
{mihail}@cc.gatech.edu

² Computer Science Department,
Berkeley, CA
{christos}@cs.berkeley.edu

Abstract. We show that the largest eigenvalues of graphs whose highest
degrees are Zipf-like distributed with slope $\alpha$ are distributed according
to a power law with slope $\alpha/2$. This follows as a direct and almost certain
corollary of the degree power law. Our result has implications for the
singular value decomposition method in information retrieval.

1 Introduction

There has been a recent surge of interest in graphs whose degrees have very
skewed distributions, with the $i$th largest degree of the graph about $ci^{-\alpha}$, for
some positive constants $c$ and $\alpha$. Such distributions are called Zipf-like distributions (the Zipf distribution being the one with $\alpha = 1$) or power laws. In contrast, the degrees of random graphs in the traditional $G_{n,p}$ model [12] are, by
the law of large numbers, exponentially distributed around the mean. It had
been observed for some time that the graph of documents and hyperlinks in the
worldwide web follow such degree distributions; in fact, there are several papers
proposing plausible models (based on “preferential attachment” [5,6,2,3,9], or
“copying”[21]) for explaining, with varying degrees of persuasiveness and rigor,
this phenomenon.

More recently, in [14] it was pointed out that the Internet graph (both the
graph of the routers and that of the autonomous systems) also has degrees that
are power law distributed, with an exponent $\alpha$ between .85 and .93. This created
much interest in the Internet research community, because the graph generators
used by researchers had theretofore lacked this property; generators that are
realistic in this sense have since appeared [18,23,24,26]. In [13] a theoretical
explanation of this phenomenon was proposed, in terms of a model of network
growth driven by the trade-off of two optimization criteria (connection costs and
communication delays), predicting a power law degree distribution.

Another very interesting, intriguing, and as of yet unexplained observation
in [14] is that the (twenty or so) largest eigenvalues of the Internet graph (that is, the largest eigenvalues of its adjacency matrix) are also power law distributed,
with $\alpha$ between .45 and .5. This is in line with similar observations in physics with $a = .5$ (see [15,16] where a heuristic explanation is described). In fact, all graph generators aiming to accurately simulate Internet topologies use the eigenvalue power law as a performance measure [18,23,24,26].

The distribution of the largest eigenvalues of Internet-related graphs is of additional special interest for the following reason: Spectral techniques [19,27, 4,1] based on the analysis of the largest eigenvalues and eigenvectors of the web graph have proven algorithmically successful in detecting “hidden patterns” such as semantics and clusters in the worldwide web. Is the Internet graph also amenable to such analysis?

In this note we provide a very simple, intuitive, and rigorous explanation of the eigenvalue power law phenomenon: We point out that it is a rather direct and almost certain corollary of the degree power law. In particular, we consider a random graph model whose degrees are, in expectation, $d_1,\ldots,d_n$ and show that, if these degrees are power-law distributed, then, with high probability, the few largest eigenvalues of the graph are close to $\sqrt{d_1},\sqrt{d_2},\ldots$ —and therefore follow a power law with exponent half of that of the degrees. This is in good agreement with the findings of [14], where the eigenvalue exponent is a little larger than half that of the degree exponent.

There is a negative implication of our result: By being essentially determined by the largest degrees (a very “local” aspect of a graph), the largest eigenvalues are unlikely to be helpful in analyzing and understanding the structure of the Internet topology (the corresponding eigenvectors are highly concentrated on the largest degrees). In [25] we show experimental evidence that spectral analysis of the Internet topology becomes useful only after the high degrees have been suitably normalized. A similar problem in the use of spectral methods in “term-document” contexts is known as the “term norm distribution problem” [17]; it is considered the main bottleneck in the use of spectral filtering for information retrieval. Extending our study from the context of undirected graphs (symmetric matrices) to the context of terms and documents (general matrices) is an interesting technical problem with direct practical significance.

The rest of the paper is organized as follows: In Section 2 we review some basics from algebraic graph theory and matrix perturbation. We state a first theorem that indicates the effect of high degrees on the spectrum. In Section 3 we show that for a rich class of random graphs whose high degrees are Zipf distributed follow a power law on their highest eigenvalues, almost surely. In Section 4 we discuss implications of our results for the singular value decomposition method.

2 Eigenvalues and Degrees

We begin by recalling certain basic facts from algebraic graph theory and matrix perturbation. For a symmetric graph $G$ with $n$ nodes, we denote by $\lambda_1(G) \geq \lambda_2(G) \geq \cdots \geq \lambda_n(G)$ the eigenvalues of its adjacency matrix in non-increasing order. $E$ denotes the set of edges, and $d_i$ the highest degree of the graph.
Fact 1. (See Lovász [22], pages 70-73.)
1. For any graph $G$, $|\lambda_i(G)| \leq \min\{|d_i, \sqrt{|E|}\}$.
2. If $G$ is a star with $n - 1$ leaves, then $\lambda_1(G) = \sqrt{n-1}$, $\lambda_n(G) = -\sqrt{n-1}$ and $\lambda_i(G) = 0, i = 2, \ldots, n-1$.
3. The multiset of the eigenvalues of a graph $G$ is the union of the eigenvalue multisets of its connected components.

Now let $A$ and $B$ be symmetric $n \times n$ matrices and let $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$ and $\lambda_1(B) \geq \lambda_2(B) \geq \cdots \geq \lambda_n(B)$ be their eigenvalues in non increasing order.

Fact 2. (See Wilkinson [30], page 101.)
$\lambda_i(A) + \lambda_n(B) \leq \lambda_i(A + B) \leq \lambda_i(A) + \lambda_i(B)$.

Now the following theorem is immediate:

Theorem 1. Suppose that an undirected graph $G$ can be decomposed into

- Vertex disjoint stars $S_i$ with degrees $d_i, i = 1, \ldots, k$.
- Vertex disjoint components $G_j$ with corresponding maximum degrees $d(G_j)$ and number of edges $e(G_j)$, such that $\min\{d(G_j), \sqrt{e(G_j)}\} = o(d_k)$, $j = 1, \ldots, m$. In addition all the components $G_j$ are disjoint from all the stars $S_i$.
- A graph $H$ with maximum degrees $d$ and $E$ edges such that $\min\{d, \sqrt{E}\} = o(d_k)$, where $H$ can have arbitrary intersections with the $S_i$'s and the $G_j$'s.

Then the largest eigenvalues of $G$ are $\sqrt{d_i}(1 - o(1)) \leq \lambda_i \sqrt{d_i}(1 + o(1)), i = 1, \ldots, k$.

Remark 1: It is clear that the spectrum of $G$ is dominated by the spectrum of the highest degree stars. It is worth noticing how much information this dominance can hide. In particular, $H$ could be any sparse graph: connected, disconnected, with or without clusters, a tree, an expander. However, we would not be able to retrieve the structure of $H$ from the spectrum of $G$. If $G$ was the topology of the Internet, $H$ could be the network backbone, and yet, all information about this structure would be lost. Indeed, in experiment, we have been able to decompose the Internet topology analyzed in [14] precisely along the lines of Theorem 1. The mere numbers are striking: For highest degree vertices in November 2000 $d(UUNET)=2034$, $d(Sprint)=1079$, $d(C&WUSA)=793$, $d(AT&T)=742$, $d(BBN)=529$, $d(QWest)=483$, $d(AboveNet)=405$, $d(Verio)=363$, $d(BusInte)=347$, $d(GlobCros)=311$ and $d(Level3)=274$, the highest eigenvalues squared were $\lambda_1^2 = 3113$, $\lambda_2^2 = 1135$, $\lambda_3^2 = 787$, $\lambda_4^2 = 676$, $\lambda_5^2 = 590$, $\lambda_6^2 = 515$, $\lambda_7^2 = 424$, $\lambda_8^2 = 395$, $\lambda_9^2 = 289$, $\lambda_{10}^2 = 277$, $\lambda_{11}^2 = 268$.

Remark 2: A technical statement analogous to Theorem 1 can be made about the stability of eigenvectors that correspond to largest eigenvalues (along the lines of Stewart [29]). As expected, the statement is that these eigenvectors of $G$ are very “close” to the eigenvectors of the stars, and are hence highly concentrated on the vertices with the highest degrees.
3 Random Graphs

We next consider a distribution of graphs with prescribed degrees. Let \( d = (d_1, d_2, \ldots, d_n) \) be a vector of integers between 0 and \( n \), in decreasing order. Denote \( \sum_{i=1}^n d_i \) by \( D_1 \), and assume that \( d_1^2 \leq D_n \). Define now \( G_n(d) \) to be the distribution of graphs with \( n \) nodes generated by the following experiment: Edges are added by independent draws, where, for \( i, j = 1, \ldots, n \) the probability that the edge \([i, j]\) is added is \( \frac{d_id_j}{D} \). Notice that we allow self-loops, an analytical convenience that does not affect the highest degrees much. Notice also that, by definition, node \( i \) has degree \( d_i \) in expectation. This random graph model has been also considered in [8] and is known to have robust connectivity properties. Our main result is the following:

**Theorem 2.** For any constant \( \gamma \), with \( 0 < \gamma < 1 \), for any constant \( \alpha \), with \( \frac{1}{2} < \alpha < 1 \), for any constant \( \beta \), with \( 0 < \beta < \frac{1}{2\alpha} \), and for any positive integer \( c \), if \( d = (d_1, d_2, \ldots, d_n) \), with

\[
d_i^2 = D_n = \Theta(n^{1+\gamma})
\]

and

\[
d_i = \frac{d_i}{i^\alpha}, \quad \text{for } i = 1, \ldots, k = \Theta(n^\beta),
\]

then, for any constant \( \beta' \), with \( 0 < \beta' < \beta \), the eigenvalues of \( G_n(d) \) satisfy

\[
\sqrt{d_i(1-o(1))} \leq \lambda_i \leq \sqrt{d_i(1+o(1))}, \quad \text{for } i = 1, \ldots, k' = \Theta(n^{\beta'}),
\]

with probability at least \( 1 - O(n^{-c}) \), for large enough \( n \) (n \( \geq \) \( n_0(\alpha, \gamma, c) \)).

**Proof.** We decompose \( G \) into the following graphs:

- \( G_1 \) is a union of vertex disjoint stars \( S_1, \ldots, S_k \) where, for \( i = 1, \ldots, k \), \( S_i \) has node \( i \) as its center and leaves those nodes from among \( k+1, \ldots, n \) which are adjacent to \( i \) and not adjacent to any node in \( \{1, \ldots, i-1\} \).
- \( G_1^* \) contains all edges of \( G \) with one endpoint in \( \{1, \ldots, k\} \) and the other in \( \{k+1, \ldots, n\} \), except those in \( G_1 \).
- \( G_2 \) is the subgraph of \( G \) induced by \( \{1, \ldots, k\} \).
- \( G_3 \) is the subgraph of \( G \) induced by \( \{k, \ldots, n\} \).

We will show that the spectrum of \( G_1 \) dominates and that each star \( S_i \) has degree very close to its expectation \( d_i \). Let \( s_i \) be the expected degree of \( S_i \) in \( G_1 \). To get a lowerbound, define \( F_i \) as the subset of vertices \( \{k+1, \ldots, n\} \) not adjacent to \( \{1, \ldots, i-1\} \) and notice:

\[
s_i = \sum_{i=k+1}^n \frac{d_i d_{i'}}{D_i} - \sum_{i \in F_i} \frac{d_i d_{i'}}{D_i} \geq d_i \sum_{i=k+1}^n \frac{d_i}{D_i} - \frac{d_i}{D_i} \sum_{i \in F_i} d_i
\]

\[
\geq d_i \sum_{i=k+1}^n \frac{d_i}{D_i} - \frac{d_i \|F_i\|}{D_i} \quad \text{ (4)}
\]
In the above expression we need to argue about the quantities \( \sum_{l=k+1}^{n} \frac{d_l}{n^{\alpha}} \) and \( E[|F_i|] \). For the first sum first notice:

\[
\sum_{j=1}^{k} d_j = d_1 \sum_{j=1}^{k} j^{-\alpha} \\
\simeq d_1 \frac{k^{1-\alpha}}{1-\alpha}, \text{ by approx with an integral} \\
= \frac{n^{1+\gamma} n^{\frac{\beta(1-\alpha)}{1-\alpha}}}{n^{1-\frac{\beta}{2\alpha}}}, \text{ by equations (1) and (2)} \\
< n^{1+\gamma} \frac{n^{\frac{\beta(1-\alpha)}{1-\alpha}}}{n^{1-\frac{\beta}{2\alpha}}}, \text{ which for } \beta < \frac{1}{2\alpha} \text{ becomes} \\
= \Theta(D_n) \frac{n^{\frac{\beta(1-\alpha)}{1-\alpha}}}{n^{1-\frac{\beta}{2\alpha}}}, \text{ with } \alpha > \frac{1}{\gamma}.
\]

Now (1) and (5) imply

\[
\sum_{l=k+1}^{n} d_l \simeq D_n.
\]

It can be seen that equation (6) above can be satisfied provided the average degree of nodes \( k+1 \) through \( n \) is \( \Omega(D_n/n) \). But the maximum expected degree of these nodes is \( d_k \), which implies that \( nd_k = \Omega(D_n) \). From equations (1) and (2) this is equivalent to \( n \cdot n^{\frac{\beta(1-\alpha)}{1-\alpha}} \cdot n^{\beta \alpha} = \Omega(n^{1+\gamma}) \), which is indeed satisfied for \( \beta \) as in the statement of Theorem 2.

For \( E[|F_i|] \) we have:

\[
E[|F_i|] = \sum_{j=k}^{n} \sum_{l=k+1}^{n} \frac{d_j d_l}{n^{\alpha}} \\
= \sum_{j=1}^{n} d_j j^{-\alpha} \sum_{l=k+1}^{n} \frac{d_l}{n^{\alpha}} \\
\simeq d_1 \frac{k^{1-\alpha}}{1-\alpha}, \text{ by approx with an integral and equation (6)}. \\
\leq d_k \frac{k^{1-\alpha}}{1-\alpha}.
\]

Now combining (4), (6) and (7) we get:

\[
s_i \geq d_i (1 - \frac{k d_1 k^{1-\alpha}}{P_n (1-\alpha)}) \\
= d_i (1 - \frac{k d_1 k^{1-\alpha} n^{\beta \alpha}}{P_n (1-\alpha)}) \text{, by substitution} \\
= d_i (1 - n^{\beta (1-2\alpha)}) \text{, with } \alpha > \frac{1}{\gamma}.
\]

Combining 8 with the obvious upper bound we get

\[
d_i (1 - n^{\beta (1-2\alpha)}) \leq s_i \leq d_i, \text{ with } \alpha > \frac{1}{\gamma}.
\]

To argue about sharp concentration of the degrees of the \( S_i \)'s around their means we will use the standard Chernoff bounds for small probabilities of success [28], Lecture 4]: For independent random variables \( X_1, \ldots, X_N \), such that such that \( \Pr[X_i = 1] = p_i \) and \( \Pr[X_i = 0] = 1 - p_i \), and where \( p = (\sum_{i=1}^{N} p_i)/N \):

\[
\Pr\left[ \left| \sum_{i=1}^{N} X_i - pN \right| > s \right] < e^{-\frac{s^2}{2N p(1-p)}} \approx \frac{\alpha^{s^2/2(N p(1-p))^{1+1}}}{\alpha^{s^2/2(N p(1-p))^{1+1}}}
\]

(10)
It can be readily checked from (9) and (10) that for some constant \( c' \), the actual degrees \( \hat{s}_i \) are concentrated as follows:

\[
d_i - \sqrt{c'd_i \log n} \leq \hat{s}_i \leq d_i + \sqrt{c'd_i \log n}
\]  

(11)

and the probability that (11) fails even for one \( i = 1, \ldots, k \) is at most \( n^{-\epsilon}/4 \). Now Fact ?? implies that the largest eigenvalues of \( G_1 \) are

\[
\sqrt{d_i(1 - o(1))} \leq \lambda_i(G_1) \leq \sqrt{d_i(1 + o(1))}, \quad i = 1, \ldots, k
\]  

(12)

and the probability that (12) fails even for one \( i = 1, \ldots, k \) is at most \( n^{-\epsilon}/4 \).

Let \( m_i \) be the expected degree of vertex \( i \) in the graph \( G'_1 \), \( i = 1, \ldots, k = n^3 \). Then using the calculations of (7) and straightforward substitutions we get:

\[
m_i = \sum_{j \in F_i} \frac{d_j}{D_j} \\
= \frac{d_k}{D_k} \sum_{j \in F_i} d_j \\
\leq d_i \cdot \frac{d_k k^{|F_i|}}{D_k} \\
\approx d_i \cdot \frac{n^{-d_i(1-\alpha)}}{n^3} \\
\leq d_i \cdot \frac{e^{\beta(1-\alpha)}}{\beta(1-\alpha)} , \text{ with } \alpha > \frac{1}{4}.
\]  

(13)

For the expected degree of vertex \( i \) in the graph \( G'_1 \) when \( i = k+1, \ldots, n \) we have the obvious bound \( d_i \leq d_k \). This together with (13) and the Chernoff bound (10) suggest that, for some constant \( c'' \) all actual degrees \( \hat{t}_i \) of \( G'_1 \) satisfy:

\[
\hat{t}_i \leq d_k + \sqrt{c''d_k \log n}
\]  

(14)

and the probability that (14) fails even for one \( i = 1, \ldots, n \) is at most \( n^{-\epsilon}/4 \).

The total number of edges for the graph \( G_2 \) is:

\[
\sum_{i=1}^{k} \sum_{j=1}^{k} \frac{d_i d_j}{D_i D_j} = \frac{d_k^2}{D_k} \sum_{i=1}^{k} i^{-\alpha} \sum_{j=1}^{k} j^{-\alpha} \\
\leq \frac{d_k^2 k^2(1-\alpha)}{\beta(1-\alpha)} \\
= n^{2\beta(1-\alpha)}
\]  

(15)

The above together with (10) suggest that, for some constant \( c''' \) the actual total number of edges \( e(G_2) \) satisfy

\[
\Pr[e(G_2) > n^{2\beta(1-\alpha)} + \sqrt{c'''n^{2\beta(1-\alpha)} \log n}] < n^{-\epsilon}/4.
\]  

(16)

For the graph \( G_3 \) we have the obvious bound that all its degrees are in expectations bounded by \( d_k \), and hence are at most \( d_k + \sqrt{c''d_k \log n} \) with probability as in (14). Combining this with (14), (16) and Fact ?? we get that the largest eigenvalues of each one of the graphs \( G'_1 \), \( G_2 \) and \( G_3 \) are

\[
\lambda_i(G'_1), \lambda_i(G_2), \lambda_i(G_3) \leq \sqrt{d_k(1 + o(d_k))}, \quad i = 1, \ldots, k
\]  

(17)
and the probability that (17) fails even for one $i = 1, \ldots, k$ is at most $3n^{-c}/4$. We may now combine (12) and (17) and see that, for any $\beta' < \beta$, we have
\[ d_k = o(d_i), \quad i = 1, \ldots, k' = n^{\beta'} \]
hence the statement of the Theorem follows.

Remark: We stated our result for the case in which the highest $d_i$’s follow an exact power law; obviously, essentially the same conclusion holds if the degrees follow a less precise law (e.g., if the degrees are within constant multiples of the bounds). Finally, the $d^2 \leq D_n$ assumption is useful for keeping the $G_n(d)$ model simple; unfortunately, it does not hold for the Internet topology. However, the Internet, as measured in [14], does satisfy the assumption, if its few (5 or 6) highest-degree nodes are removed. These high-degree nodes do not affect the other degrees much, and do not harm our argument, no matter how adversely they may be connected.

4 Implication on SVD method for Information Retrieval

Spectral filtering and, in particular the singular value decomposition (SVD) method is repeatedly invoked in information retrieval and datamining. It has also been amenable to theoretical analysis and has yielded a remarkable set of elegant algorithmic tools [19,27,4,1]. However, in practice, SVD is weakened by the so-called “term norm distribution problem”: this arises when terms are used in frequencies disproportionately higher than their relative significance, and several heuristics (so-called “inverse frequency normalizations”) are known, however, none of them is known to perform adequately in theory or in practice (see [17] for a nice exposition). The term norm distribution problem appears very similar to the problem of high degrees that we treated here. It would be interesting to study the term norm distribution problem in a theoretical framework and quantify the proposed heuristics to overcome it. For the Internet topology of [14], in practice we solved the problem of high degrees by pruning small ISP’s (leaves and a few more nodes) [25]. However, we do not have a formal framework for this method, and we do not know how it would extend in the case of term-documents or directed graphs.

The effectiveness of several of the SVD-based algorithms [4,1] requires that the underlying space has “low rank”, that is, a relatively small number of significant eigenvalues. Power laws on the statistics of these spaces, including eigenvalue power laws, have been observed [7,20,10] and are quoted as evidence that the involved spaces are indeed low rank and hence spectral methods should be efficient. In view of the fact that the corresponding eigenvalue power law on the Internet topology was essentially a restatement of the high degrees and thus revealing no “hidden” semantics (see Remark 2 in Section 2), it is intriguing to understand what kind of information the corresponding power laws on the spectra of term-document spaces convey (or hide...
Acknowledgements

The first author was supported by NSF under grant CCR-9732746 and by a Georgia Tech Edenfield Faculty Fellowship. The second author was supported by NSF under grant CCR-0121555.

References


