

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 28 (2007)
Indiana University

**Power Law Degree Distributions Can Fit Averages of Non-Power
Law Distributions¹**

Thomas M. Gruenfelder and Shane T. Mueller²

*Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405*

¹ This research was supported by NIH Grant DC00111.

² Now a Senior Scientist at Klein Associates, a Division of ARA Inc., Fairborn, OH.

Power Law Degree Distributions Can Fit Averages of Non-Power Law Distributions

Abstract. Many complex systems have been modeled as networks. Examples of such systems that are of interest to cognitive psychologists include the mental lexicon for spoken word recognition and semantic memory. A frequent finding in such studies is that the frequency distribution of the number of connections for each node in the network follows a power law. This finding has been interpreted to mean that the network grows through a process similar to preferential attachment: when a node is added to the network, it attaches to other nodes with a probability proportional to the number of connections those other nodes already have. Power-law degree distributions, however, may also well describe degree distributions that result when averaging across multiple individual degree distributions, none of which follows a power-law. Further, each of these individual distributions may reflect a random growth process rather than the more systematic process suggested by preferential attachment.

Introduction

There has been a lot of interest over the past 10 years in using the tools of graph theory to model complex networks (see Albert & Barabási, 2002, for a review). In such studies, entities are represented as nodes or vertices, and an edge or link is placed between two nodes if some pre-determined relation exists between them. For example, in a model of the World Wide Web, each site on the web can be represented by a node. A link is placed between two nodes if the site represented by one such node has a (web) link to the site represented by the other node (see, for example, Pastor-Satorras & Vespignani, 2004).

A common finding that has emerged from these studies is that many of these real-world networks have a small-world (Watts & Strogatz, 1998), “scale-free” (Barabási & Albert, 1999) structure. In a small-world network, the mean shortest path between any two nodes in the network is relatively short. The shortest path refers to the smallest number of links that must be traversed to get from one node to the other. In particular, in a small-world structure, the mean shortest path length grows more slowly than the number of nodes. It is similar to the mean shortest path length that occurs in a random graph, where links are placed between nodes randomly. Lattice networks, in contrast, where each node connects to only a very few near neighbors, have a much longer mean shortest path length. Small-world networks are also characterized by a mean clustering coefficient (CC) larger than that expected by chance. A node’s CC is the probability that two nodes with a link to that node also have a link to one another. When we say that the CC is higher than the CC expected by chance, we mean that it is higher than what would be expected in a comparably sized network (a network with the same number of nodes and the same total number of links) in which links between nodes have been placed at random. In brief, a small-world network is a network with a relatively small mean shortest path length between any two nodes and a CC that is much higher than that expected in a random graph.

In a scale-free network, the degree distribution follows a power law, $N(k) \sim k^{-\gamma}$, where $N(k)$ is the degree distribution, k is the degree, and γ , the distribution’s parameter, is typically between 2 and 3. A node’s degree, k , refers to the number of links going into or coming out of that node. The degree distribution is simply the frequency distribution of the degree across all nodes in the network. When plotted on log-log coordinates, the slope of a power-law degree distribution, of course, is a straight line with a slope of $-\gamma$. (For this reason, we sometimes refer to γ as the slope parameter.)

A finding as ubiquitous as that of scale-free degree distributions demands an explanation and Barabási and Albert (1999) have provided one. They showed that if a network grows continuously by adding new nodes, and the new nodes form links with existing nodes with a probability proportional to the number of links that existing node already has, then a power-law degree distribution emerges. They termed this process “preferential attachment.” Less formally, it has been referred to as the “rich get richer” principle. In the psychological literature, similar phenomena have been referred to as the Matthew principle (Stanovich, 1986; see also McClelland & Rumelhart, 1981).

Many of these recent studies of complex networks are of direct interest to cognitive psychologists. Schweickert (in press a, b), for example, analyzed the co-occurrence of characters in a given person’s dream. Characters were represented as nodes and a link was placed between two nodes if those two characters had appeared in the same dream. Schweickert found that the networks for all three of the dreamers he analyzed showed a small-world structure. Two of the three also showed a power law degree distribution, at least at higher degrees (Schweickert did not use the term scale-free.). Given the similarity of these structures to social networks, he argued that analysis of the characters appearing in a person’s dream might be a reliable method for determining that person’s social network.

In another recent study, Steyvers and Tenenbaum (2005) analyzed three types of semantic networks, in which the nodes represented words. The first was based on the Nelson, McEvoy, and Scheiber (1999) word association norms. A link was placed between two words if one of those words had been produced as an associate of the other by at least two participants in the Nelson et al. norms. The second network was based on Roget’s thesaurus (Roget, 1911). In this network, a link was placed between two words if they shared at least one category in the thesaurus in common. The third network was based on WordNet (Fellbaum, 1998; Miller, 1995) in which word-form nodes are connected to word-meaning nodes and word-meaning nodes, in turn, are connected to one another based upon the relation between those meanings (such as antonymy (LOVE and HATE), hypernymy (A ROBIN is a BIRD), and meronymy (A ROBIN has WINGS)).

All three networks showed a small-world, scale-free structure. Steyvers and Tenenbaum (2005) noted, however, that other aspects of their results were inconsistent with growth through preferential attachment and proposed a somewhat different model to explain their results. In the Steyvers and Tenenbaum model, at each point in time, a node is chosen with a probability proportional to its number of links for differentiation. If node i is chosen for differentiation, then a new node is added to the network and connected to M randomly chosen nodes that are already neighbors (i.e., have a link to) node i . This model is not so much a “rich get richer” model as a “rich beget rich” model.

Ferrer and Solé (2001) built a network of words based upon their co-occurrence in sentences. In particular, they placed a link between any two words that occurred within two words of each other in a sentence in their corpus with a probability higher than that expected by chance. Their network showed a small-world structure. Its degree distribution showed two distinct power-law regions, one covering smaller degrees and a second, with a somewhat steeper slope, covering higher degrees. They interpreted this property as being consistent with the notion that the network of words in sentences is scale-free.

Soares, Corso, and Lucena (2001) modeled the syllabic structure of the Portuguese language. Each node in their network represented a word in the Portuguese language. A link was placed between two nodes if those two syllables co-occurred in the same word. Soares et al.

found that their network followed a small-world, scale-free structure. They interpreted this finding to mean that the Portuguese language evolved in a non-random manner, i.e., new words were added to the language through a process akin to preferential attachment.

Finally, Vitevitch (in press) and Gruenenfelder and Pisoni (2005) used the Hoosier Mental Lexicon data base (Nusbaum, Pisoni, & Davis, 1984) to model the mental lexicon of spoken word representations. In both of these studies, spoken word forms were represented as nodes and a link was placed between two nodes if the word represented by one could be changed into the word represented by the second by the deletion, addition, or substitution of a single phoneme (Greenberg & Jenkins, 1964; Landauer & Streeter, 1973; Luce & Pisoni, 1998). (We refer to this rule for defining lexical neighbors as the DAS rule, for Deletion, Addition, or Substitution.) Both studies found that the mental lexicon, modeled in this way, showed a small-world structure. In addition, both studies found that a power law did a good job describing the degree distribution. Vitevitch, for the subset of words that he studied, found that an exponential fit the degree distribution somewhat better than did a power function. Gruenenfelder and Pisoni found that which function provided a better fit was determined by details of how the fit was made (see below).

Gruenenfelder and Pisoni (2005) pointed out that a word's degree is confounded with its length, as measured by the number of phonemes. Shorter words have higher degrees, i.e., have more neighbors, as determined by the DAS rule, than longer words. Further, given that words are constructed from a relatively small set of basic elements or particles, i.e., phonemes, this fact has to be the case. When degree distributions were examined for particular classes of words, as determined by their length, no evidence for power-law degree distributions was evident. The power-law degree distribution emerged only when data were averaged across these multiple, non-power-law distributions. Figure 1 shows the degree distribution, collapsed across word length, reported by Gruenenfelder and Pisoni. The figure is on a log-log plot and hence a power function would be indicated by a straight line. Visual inspection of the figure does in fact suggest that the distribution is well fit by a straight line (with the exception of the last four data points, where the function "falls off the cliff"). A regression analysis showed that in fact the distribution is reasonably well fit by a straight line, $R^2 = .79$. (Incidentally, the fit of an exponential to the overall degree distribution is even better: $R^2 = .89$. If the last four data points to the right, where the function falls off the cliff, and which are inconsistent with both a power law and exponential function, are dropped from the regression analysis, the R^2 for the power law fit increases to .96; the R^2 for the exponential fit increases to only .91. Hence, the better fit of the exponential to the overall distribution seems due to its ability to better describe what are clearly contradictory data points.)

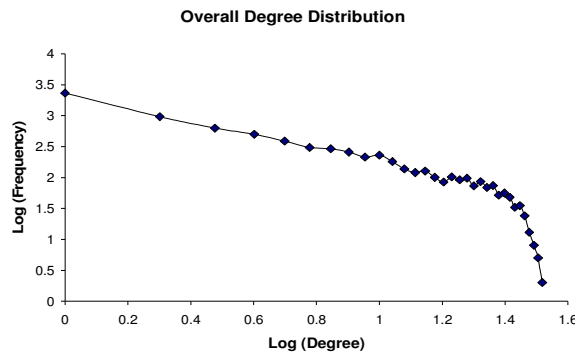


Figure 1. Overall degree distribution of neighbors, collapsed across word length, for the word corpus analyzed by Gruenenfelder and Pisoni (2005).

Figure 2 shows the degree distribution broken down by word length. None of the degree distributions in Figure 2 provides a close match to a (downward trending) power function, particularly at intermediate word lengths (3, 4, 5, and 6), which account for the majority of words. The composite power law degree distribution emerges only when the data are averaged across word lengths.

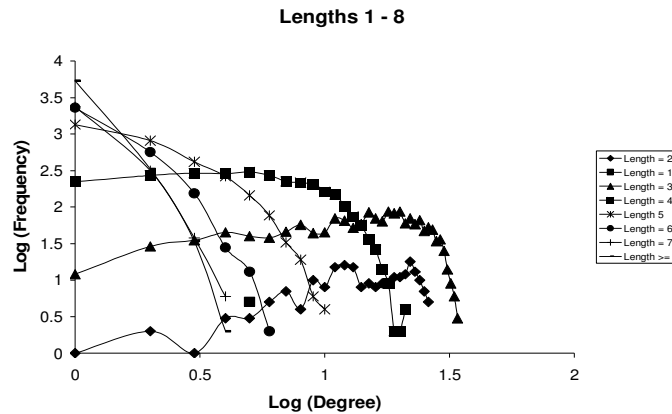


Figure 2. Degree distribution of neighbors, broken down by word length in phonemes, for the word corpus analyzed by Gruenenfelder and Pisoni (2005).

Gruenenfelder and Pisoni's (2005) interpretation is one example of a more general phenomenon that has been observed elsewhere, such as in the memory (Anderson, 2001; see also Brown & Heathcote, 2003a, 2003b) and skill-learning literatures (Newell & Rosenbloom, 1981). A power-law function can frequently well describe the result when data are averaged across a number of underlying distributions, none of which itself is a power law. In fact, the power law fit is sometimes better than a fit of a function of the same form as that of those being averaged to give the composite function.

Anderson's (2001; Anderson & Tweney, 1997) work is perhaps the most pertinent here. Anderson & Tweney were interested in modeling the memory decay curve—the decline in memory performance as a function of retention interval or time, t . They assumed that memory for an item consisted of multiple memory traces, each decaying independently as an exponential, Ae^{-Bt} (A and B are constants), each with a different decay parameter, B . (Because, when the logarithm of frequency is plotted as a function degree, in our case, or time, in Anderson & Tweney's case, an exponential yields a linear function with a slope of $-B$, we sometimes refer to B as the slope parameter.) Observed memory performance is the average of these curves. This situation would apply, for example, to the case where a single retention curve is plotted based on average retention across subjects, or where a single retention curve is plotted based on average retention of a single subject for multiple different items. Anderson and Tweney found that in a small number of their simulations (5%) (arithmetic) average performance was better fit by a power law function than by an exponential. In general, power law fits improved as rate parameters varied over a larger range. In addition, when they made the additional assumption that there was also noise in the measurement of each subject's memory performance at each retention interval, power laws fit the resulting retention curves better than exponentials in 97% of the cases. In summary, Anderson and Tweney showed that power law distributions can arise when multiple exponential distributions are averaged together.

Anderson (2001) extended Anderson and Tweney's (1997) work in two important ways. First, he showed that when the component exponential curves were constrained to have a downward slope, even when noise in the measurement of the subject's performance was not assumed, power laws tended to fit the average curve better than did exponentials provided there was sufficient variance in the rate parameters of the component distributions. Second, he showed that the same result—better fit of power laws to the average distribution than of a function of the same form as the component distributions—occurred when the component distributions were range-limited linear or range-limited logarithmic functions. By range-limited, Anderson meant that the function was never permitted to go below 0, certainly a reasonable assumption when discussing memory retention. We can remember nothing but we cannot remember a negative amount of information. Without this restriction, the average of a number of linear components is, of course, another linear function and the average of a number of logarithmic components is another logarithmic function.

Anderson's work with the exponential is especially interesting to us when applying graph theoretic tools to the analysis of complex systems for two reasons. First, Watts and Strogatz (1998) showed that for random graphs (i.e., graphs with a fixed number of nodes in which a fixed number of links is placed between randomly chosen pairs of nodes) the degree distribution follows a Poisson distribution, the right hand tail of which approaches an exponential. It is precisely at higher degrees that the power law degree distribution often becomes most evident. Schweickert (in press a, b), for example, only considered degrees above the median degree. Second, Barabási and Albert (1999) showed that if the assumption that networks grow over time is retained, but the assumption of preferential attachment is replaced with the assumption that new nodes attach randomly to existing nodes (i.e., attach to each existing node with a probability proportional to the number of existing nodes), then the degree distribution is not a power law but an exponential.

Suppose that we have a complex network that is in fact composed of several sub-networks in the following sense. The network is growing over time, but there are several different processes that can result in a new node being added to the network. The existing nodes that a new node connects to are random (i.e., a new node's probability of connecting to an existing node is proportional to the number of nodes in the network), but the constant of proportionality differs from one process to another. In this case, the observed network is the average of the several non-observed "sub-networks," and its degree distribution is the average of the exponential degree distributions of the several underlying "sub-networks." Anderson's work indicates that this aggregate degree distribution is likely to follow a power law even though all the underlying distributions are exponential. There is certainly no guarantee that the observed distribution in the composite network is of the same form as the distribution in each of the underlying sub-networks (cf. Estes, 1956).

To summarize, a power law degree distribution may result because a network grows via preferential attachment. It may also result because the network reflects several component networks, each of which grows randomly and each of which has an exponential (or perhaps even some other) degree distribution.

Are the recent studies we cited above vulnerable to this ambiguous interpretation? Could the power law degree distributions that they observe simply be the result of averaging across "random" processes rather than reflecting some fundamental property of the underlying network, such as growth through preferential attachment? The possibility certainly cannot be ruled out. The power law degree distribution for the spoken word lexicon observed by Gruenenfelder and Pisoni (2005) (and possibly the degree distribution observed by Vitevitch, in press) does seem to be the

result of averaging across words of different lengths. Steyvers and Tenenbaum's (2005) network based on association norms averages across, amongst other things, subjects. Their thesaurus based network averages across words of different syntactic classes (verbs, nouns, adjectives). Their WordNet based network averages across different semantic relations.³ Ferrer and Solé's (2001) network involved averaging over different types of documents. Soares et al. (2001) averaged across words of different lengths, as measured by number of syllables. Schweickert's (in press a; in press b) work seems less affected by this potential ambiguity, as he built separate networks for each of his dreamers. However, he did not report goodness of fit measures to exponential degree distributions.

The present study was carried out to investigate whether the occurrence of power law degree distributions could be the result, in at least some graph theoretic studies, of averaging across multiple exponential distributions, each generated by a random process, rather than the result of a more systematic growth process, such as preferential attachment. We built a simulation that created networks by adding a node to that network at each time step, t . That new node was then connected to each existing node in the network with probability of \underline{r}/N_t , where N_t is the total number of nodes in the network at time step t , and \underline{r} , referred to above as the constant of proportionality, is a parameter that we varied across simulations. The simulation continued until some pre-specified maximum number of nodes, N , had been created. We then computed the degree distribution for that network. Based on the work of Barabási and Albert (1999) we expected (and found) that each individual network created in that fashion would yield an exponential degree distribution.

We then "simulated" a composite network by summing the degree distribution of multiples of these individual networks. We then fit the degree distribution of this composite network to linear, exponential, and power functions to determine which function best described the degree distribution of the composite network.

Simulation I

In their simulations, Anderson (2001) and Anderson and Tweney (1997) directly manipulated the slope parameter of the underlying exponential functions that they were simulating. They found that for a power law to best fit the composite function, there needed to be sufficient variability in the slope parameters of the underlying exponentials. The present study does not directly simulate an exponential degree distribution. Instead, it simulates a process that, as it turns out, produces an exponential degree distribution. The slope parameter of that exponential is affected by both N , the size of the network in number of nodes, and \underline{r} , the probability that a new node will attach to any already existing node in that network. Simulation I investigated more precisely how \underline{r} affected the slope parameter for individual exponential degree distributions for different sized networks. The purpose was to allow, in the main set of simulations, choosing values of \underline{r} such that there would be sufficient variation in the slope parameters for good power law fits to the composite degree distributions to emerge.

As already mentioned, the present simulation built a network by adding a node at each time step. That node was then connected to each already existing node in the network with probability \underline{r}/N_t , where N_t is the number of nodes in the network at time step t . The simulation

³ To be fair, Steyvers and Tenenbaum (2005) noted discrepancies between their model and the preferential attachment model. They proposed a revised model, sketched above, to account for their results. It is not clear that a model averaging random networks could capture these additional aspects of their data.

continued until N nodes had been added to the network. Both r and N were user supplied parameters to the simulation.⁴

Method

Four different size networks were simulated in Simulation I. The four network sizes were $N = 20,000, 40,000, 80,000$ and $100,000$ total nodes. For each network size, r was varied from $.01$ to $.1$ in steps of $.01$, from $.1$ to 1.0 in steps of $.05$, and from 1 to 2 in steps of $.1$ (a total of 38 r values for each network size). Each combination of N and r was simulated once, for a total of 4×38 or 152 simulations. The degree distribution was determined for each simulation and the best fitting exponential to that degree distribution was computed using regression analysis.

Results and Discussion

Recall that for an exponential degree distribution, the logarithm of a degree's frequency is a linear function of the degree. Consequently, linear regression analysis can be used to compute the best fitting exponential to an observed degree distribution and for estimating the slope parameter of that exponential. That method was used in analyzing the present results. It does preclude including hermit nodes—nodes with no neighbors—in the analyses since the logarithm of 0 is undefined. Excluding these nodes has at worst a small effect on the results reported here. It is also the case that the literature on the analyses of degree distributions does not seem to point to a uniformly accepted way of handling these hermit nodes.

The best fitting exponential was determined for each of the 152 degree distributions computed in Simulation I. R^2 values for these fits are shown in Figure 3 as a function of r and N . Overall, exponential functions fit these distributions extremely well. The mean R^2 values for network sizes of $20,000, 40,000, 80,000$ and $100,000$ nodes were, respectively, $.962, .946, .899,$ and $.868$. Because it is already known that networks grown in this random fashion produce exponential degree distributions (see, for example, Barabási & Albert, 1997), these results are not new and were not unexpected. They do provide some confirmation that the simulation was in fact simulating the intended process.

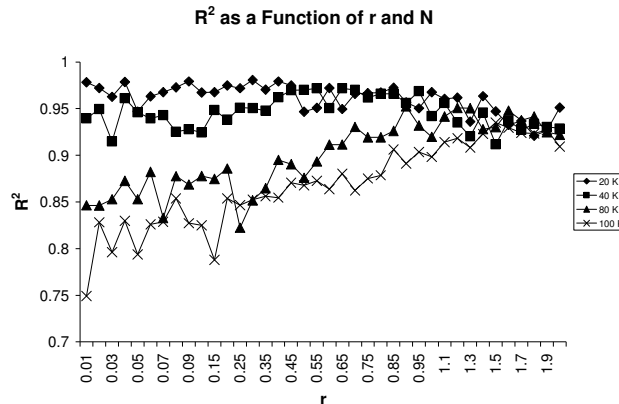


Figure 3. R^2 as a function of r . The curve parameter is N , network size.

⁴ There were two additional parameters input to the simulation as well. However, as these were not varied as part of the present study, they are not discussed here.

The fact that the fits became less good as network size grew is perhaps at least in part attributable to the fact that the larger networks had a larger number of degrees. Hence, more data points were being fit for the degree distributions from the larger networks.

The more important results concern how the slope of these best fitting exponential degree distributions changes as a function of \underline{r} and N . These results are shown in Figure 4. For small values of \underline{r} , smaller networks show smaller magnitude slopes in the degree distribution than do larger networks. The smaller networks also show more change in the slope as \underline{r} varies. Above values of $\underline{r} \approx 1.0$, network size does not appear to have much influence on the slope of the degree distribution. Similarly, the effect of \underline{r} itself on the slope appears to be much larger at values of \underline{r} less than approximately 1.0. Above that value, the slope changes much more gradually with \underline{r} . Accordingly, in our main simulation, we worked with relatively small networks ($N = 20,000$) and with mean values of \underline{r} under 1.0.

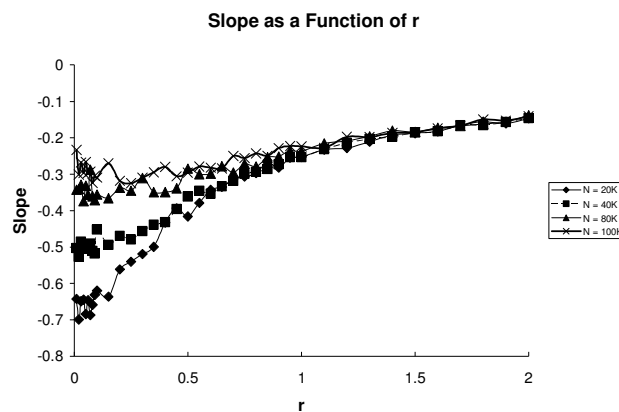


Figure 4. Slope of the predicted degree distributions as a function of \underline{r} and N .

Simulation II

Simulation II comprised the main simulation of the present study. Its purpose was to determine how well the degree distribution of a composite network, composed of a number of individual random networks, each of which would be expected to have an exponential degree distribution, could be fit with a power law function. Based on the results of Simulation I, all individual networks created in Simulation II were relatively small, consisting of 20,000 nodes. A number of composite networks were built as part of Simulation II. These networks differed from one another in the mean value of the \underline{r} parameter used to create the individual networks and its standard deviation.

Method

A total of 70 composite networks were simulated by factorially combining 7 mean values of the \underline{r} parameter of the underlying network (.15 through and including .75 in steps of .1) with 10 values of the standard deviation of \underline{r} (.1 through and including 1.0 in steps of .1). Each composite network was formed by summing together 20 individual networks, each of which consisted of 20,000 nodes. Each of the individual networks was formed by randomly sampling a value of \underline{r} from a normal distribution with a mean and standard deviation corresponding to the mean and standard deviation of the corresponding cell in the design. The sampling was done with the restriction that \underline{r} could not take on negative values. If a negative value of \underline{r} was sampled it was discarded and the distribution re-sampled. The result of this sampling procedure is that the actual

mean values of \bar{r} , especially for cells with low means and high standard deviations, was somewhat larger than the nominal means, sometimes greatly so. For example, the actual mean value of \bar{r} for the individual networks sampled from a distribution with a mean of .15 and standard deviation of .8 was .48.

Results

The degree distribution of each of the 70 composite networks was calculated. The linear, exponential, and power law distributions that best fit these observed distributions were then determined using linear regression analyses. (The linear degree distribution is linear when frequency is plotted as a function of degree. The exponential degree distribution is linear when the logarithm of frequency is plotted as a function of degree. And the power law degree distribution is linear when the logarithm of frequency is plotted as a function of the logarithm of the degree.) Because the logarithm of 0 is undefined, nodes with degree 0 (i.e., with no edges) could not be included when fitting the power law degree distribution. These nodes were also excluded when fitting the linear and exponential distributions so that all three functions would be fitting the same data points. R^2 was used as the measure of goodness of fit.

Tables 1, 2, and 3 show the R^2 values for the best fitting linear, exponential, and power law degree distributions, respectively, as a function of the mean value of \bar{r} and its standard deviation. As expected, linear degree distributions do not provide a very good fit to the observed degree distributions. Across the 70 composite networks, the mean R^2 for the best fitting linear degree distribution was .446. Exponential distributions, in contrast, provide excellent fits to the observed degree distributions. Across the 70 composite networks, the mean R^2 for the best fitting exponential degree distribution was .989. The goodness of fit of the exponentials did not vary much as \bar{r} or its standard deviation varied, at least in part because of ceiling effects. Power law degree distributions produced what might be best called good fits to the observed degree distributions. Across the 70 composite networks, the mean R^2 for the best fitting power law degree distribution was .873. The goodness of fit of the power law did not vary much as the standard deviation of \bar{r} varied. There is, however, a trend to somewhat worse fits as the mean value of \bar{r} increased. Although the data in Table 3 indicate reasonable power law fits to the composite degree distributions, a comparison of Tables 2 and 3 clearly show that the exponential fits are superior to the power law fits. In none of the 70 composite networks did a power law better fit the degree distribution than did the exponential.

	Standard Deviation									
Mean	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
.15	.517	.436	.410	.434	.456	.458	.428	.354	.323	.454
.25	.555	.436	.493	.414	.453	.438	.419	.397	.442	.329
.35	.552	.490	.484	.430	.482	.491	.456	.325	.365	.376
.45	.556	.476	.455	.427	.441	.414	.409	.439	.423	.470
.55	.508	.524	.433	.419	.462	.475	.378	.428	.409	.342
.65	.555	.511	.516	.480	.424	.440	.441	.434	.481	.371
.75	.517	.451	.468	.530	.438	.454	.429	.488	.438	.384

Table 1. Goodness of fit (R^2) measures for linear fits to the observed degree distributions for the 70 composite networks, as a function of \bar{r} and its standard deviation.

	Standard Deviation									
Mean	.1	.2	.3	.4	.5	.6	.7	8	.9	1.0
.15	.996	.992	.972	.988	.995	.995	.997	.987	.983	.991
.25	.996	.992	.998	.993	.994	.994	.996	.977	.997	.973
.35	.995	.999	.990	.996	.992	.995	.997	.961	.991	.976
.45	.997	.986	.994	.969	.991	.987	.994	.988	.979	.993
.55	.994	.995	.993	.976	.992	.993	.957	.988	.989	.984
.65	.991	.997	.994	.992	.980	.984	.988	.988	.992	.979
.75	.992	.987	.993	.996	.988	.984	.991	.995	.988	.976

Table 2. Goodness of fit (R^2) measures for exponential fits to the observed degree distributions for the 70 composite networks, as a function of \bar{r} and its standard deviation.

	Standard Deviation									
Mean	.1	.2	.3	.4	.5	.6	.7	8	.9	1.0
.15	.901	.913	.898	.902	.877	.869	.880	.897	.897	.855
.25	.874	.897	.883	.886	.894	.866	.868	.892	.871	.902
.35	.865	.873	.861	.891	.870	.854	.865	.906	.882	.875
.45	.863	.870	.879	.866	.871	.884	.884	.859	.878	.852
.55	.875	.853	.894	.881	.872	.849	.895	.855	.880	.881
.65	.831	.860	.859	.858	.880	.864	.873	.857	.848	.856
.75	.850	.873	.871	.868	.863	.855	.869	.857	.860	.882

Table 3. Goodness of fit (R^2) measures for power law fits to the observed degree distributions for the 70 composite networks, as a function of \bar{r} and its standard deviation.

Discussion

The present study grew networks by simulating a process known to produce exponential degree distributions. It then formed composite networks by adding together several of these individual networks and examined the degree distributions of those composite networks. Based on the earlier work of others (e.g., Anderson, 2001; Anderson & Tweney, 1997), we hypothesized that the degree distributions of those composite networks might be best fit by a power law function. Contrary to that hypothesis, the degree distribution of every composite network was better fit with an exponential than with a power law. Hence, at least for the parameter space explored here, it is tempting to conclude that when evaluating the degree distributions of networks, researchers do not need to be overly concerned with the possibility of power law mimicry. Furthermore, because we used network sizes and values of \bar{r} intended to maximize the variability in the slopes of the individual networks' degree distributions, a condition Anderson and Tweney (1997) found necessary to produce power law mimicry, it may also be tempting to conclude that it is unlikely that power law mimicry will turn out to be an extensive problem when evaluating the degree distributions of complex systems.

However, power laws did provide relatively good fits to the composite networks, even though our individual networks were grown in a completely random manner, with no process akin to preferential attachment operating. Consequently, observing a good fit of a power law function to a degree distribution is not sufficient evidence that a process such as preferential attachment is operating. Minimally, the degree distribution also needs to be fit to an exponential distribution. If the exponential fits better than the power law, there is little evidence that anything

other than random processes are operating. To the extent that the power law provides a better fit than the exponential, given the results of the present simulations, increased confidence can be put into a claim that the observed network grew via a process such as preferential attachment. However, that observation alone cannot exclude the possibility that such a study had the misfortune of stumbling into some area of the parameter space where power law mimicry was operating.

Regrettably, in those studies using graph-theoretic analyses in areas of most interest to cognitive science, it does not seem to be routine practice to report fits of exponential functions to observed degree distributions. Steyvers and Tenenbaum (2005) reported that for 3 or their 4 semantic networks, a power law “almost perfectly” (p. 53) fit the observed degree distribution. The degree distribution of the fourth network “show[ed] a slight deviation from the power law form” (p. 54). (Although Steyvers and Tenenbaum seemed to conclude from this finding (along with the observed slopes of those distribution) that their networks generally showed a scale-free structure, they did reject the hypothesis of growth through preferential attachment due to other observed characteristics of their network.) They reported no quantitative measure of goodness of fit, and did not report on how well those degree distributions were fit by an exponential function. Ferrer and Solé (2001) similarly fit their degree distributions of the co-occurrence of English words with a power law but did not compare that fit to the fit of an exponential. Soares et al. (2005), in their study of the syllabic structure of Portuguese reported good fits of their degree distributions to power functions. They did not, however, report a measure of the goodness of fit, nor did they compare that fit to the fit of an exponential. They concluded that language evolves following a process similar to the preferential attachment rule.

Schweickert (in press b), in his study of the networks of characters appearing in individuals’ dreams, did report goodness of fit measures for his fits to a power law degree distribution. The R^2 values for the networks of his three dreamers were .84, .85, and .82. These values are slightly lower than most of the R^2 values we observed for the random composite networks grown in the present study. Unfortunately, Schweickert did not report corresponding goodness of fit measures for exponential fits to his degree distributions.

Vitevitch (in press) did report on the fit of both power law and exponential functions to his degree distribution. He found a better fit for the exponential than for the power function. Interestingly, he then speculated on what developmental processes might produce such a degree distribution without mentioning the simplest such process: a network that grows over time, in which newly added nodes randomly attach to existing nodes with a probability proportional to the current size of the network (Barabási & Albert, 1999). That is, he did not consider the type of random graph simulated in the present study.

The degree distribution, of course, is not the only characteristic of graphs that can differentiate different growth processes that give rise to the observable network. Steyvers and Tenenbaum (2005), for example, rejected preferential attachment as the growth mechanism underlying their observed semantic networks because the clustering coefficient in their graphs was much larger than that predicted by preferential attachment. Although we have not yet done a systematic examination, we did look at the clustering coefficient in a few of the graphs generated via our random growth process and found them to be extremely small.⁵ Soares et al. (2005), in their study of the syllabic structure of Portuguese words, and Vitevitch (in press), in his study of 6508 English words, similarly observed larger clustering coefficients than occur in a certain form

⁵ Note that these clustering coefficients were calculated from the individual sub-networks, not from the composite networks.

of random graph, different from that studied here. At the present time, however, it is unclear whether such an observation actually does rule out growth through random processes in the domains studied by these researchers. Both studies considered a domain that has what Abler (1989) refers to as a particulate structure. In such a domain, an infinite or near infinite set of structures is constructed from a much smaller set of particulate units. The Portuguese words studied by Soares et al. were created out of syllable particulates. The English words studied by Vitevitch were created out of phoneme particulates. In each case, a relatively small number of particles is used to create a relatively large number of words. In von Humboldt's words (cited in Abler), words are constructed by "mak[ing] infinite use of finite media." Words are then linked by virtue of what amounts to having particles in common. Intuitively—and a stronger statement awaits further investigation—such a structure seems likely to produce dense neighborhoods with relatively high clustering coefficients. In brief, it has not been shown that a relatively high clustering coefficient necessarily rules out random growth processes.

To summarize, power law degree distributions are frequently observed when modeling a complex system as a graph. Such a ubiquitous finding inspires a search for a common cause. A possible common cause in this case is growth through preferential attachment. However, the random growth process described in the present simulations produced degree distributions that, though better fit by an exponential, were also well fit by a power law. As a science, we need to rule out simple explanations based on random processes before moving on to the more complex explanations.

References

- Abler, W.L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social & Biological Structures*, 12, 1-13.
- Albert, R. & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47-97.
- Anderson, R.B. (2001). The power law as an emergent property. *Memory & Cognition*, 29, 1061-1068.
- Anderson, R.B. & Tweney, R. (1997). Artifactual power curves in forgetting. *Memory & Cognition*, 25, 724 - 730.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509-512.
- Brown, S. & Heathcote, A. (2003a). Bias in exponential and power function fits due to noise: Comment on Myung, Kim and Pitt. *Memory and Cognition*, 31, 656-661.
- Brown, S. & Heathcote, A. (2003b). Averaging learning curves across and within participants. *Behaviour Research Methods, Instruments & Computers*, 35, 11-21.
- Estes, W.K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, 53, 134 - 140.
- Fellbaum, C. (Ed.) (1998). *WordNet, an electronic lexical database*. Cambridge, MA: MIT Press.
- Ferrer, R. & Solé, R.V. (2001). The small world of human language. *Proceedings of the Royal Society of London B*, 268, 2261-2265.
- Greenberg, J.H. & Jenkins, J.J. (1964). Studies in the psychological correlates of the sound system of American English. *Word*, 20, 157-177.
- Gruenenfelder, T.M. & Pisoni, D.B. (2005). Modeling the mental lexicon as a complex system: Some preliminary results using graph theoretic measures. In *Research on Spoken Language Processing Progress Report No. 27* (pp. 27-27). Bloomington, IN: Speech Research Laboratory, Indiana University.

- Landauer, T.K. & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 12, 119–131.
- Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- McClelland, J.L. & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375 – 407.
- Miller, G. A. (1995). WordNet: An on-line lexical database[Special issue]. *International Journal of Lexicography*, 3(4).
- Nelson, D.L., McEvoy, C.L., & Schreiber, T.A. (1999). *The University of South Florida word association norms*. Available at <http://w3.usf.edu/FreeAssociation>
- Newell, A. & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10*, Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.
- Roget, P. M. (1911). *Roget's thesaurus of English words and phrases* (1911 ed.). Retrieved (by Steyvers & Tenenbaum) October 28, 2004 from <http://www.gutenberg.org/etext/10681>
- Schweickert, R. (in press a). The structure of semantic and phonological networks and the structure of a social network in dreams. In J.S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III*. New York: Psychology Press.
- Schweickert, R. (in press b). Properties of the organization of memory for people: Evidence from dream reports. *Psychonomic Bulletin & Review*.
- Soares, M.M., Corso, G., & Lucena, L.S. (2005). The network of syllables in Portuguese. *Physica A: Statistical Mechanics and its Applications*, 355, 678- 684.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Quarterly Journal of Reading*, 21, 360 – 407.
- Steyvers, M. & Tenenbaum, J.B. (2005). The large-scale structure of semantic networks: Statistical analysis and a model of semantic growth. *Cognitive Science*, 29, 41-78.
- Vitevitch, M.S. (in press). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech-Language-Hearing Research*.
- Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440-442.