The small world phenomenon in hybrid power law graphs

Fan Chung and Linyuan Lu[1]

University of California, San Diego

Abstract. The small world phenomenon, that consistently occurs in numerous existing networks, refers to two similar but different properties — small average distance and the clustering effect. We consider a hybrid graph model that incorporates both properties by combining a global graph and a local graph. The global graph is modeled by a random graph with a power law degree distribution, while the local graph has specified local connectivity. We will prove that the hybrid graph has average distance and diameter close to that of random graphs with the same degree distribution (under certain mild conditions). We also give a simple decomposition algorithm which, for any given (real) graph, identifies the global edges and extracts the local graph (which is uniquely determined depending only on the local connectivity). We can then apply our theoretical results for analyzing real graphs, provided the parameters of the hybrid model can be appropriately chosen.

1 Introduction

In 1967, the psychologist Stanley Milgram [32] conducted a series of experiments which led him to the well known concept captured by the phrase "Six degrees of separation". Namely, any two strangers (on the planet) are *connected* by a short chain of intermediate acquaintances of length at most six. Since then, it has been observed that many realistic networks possess the so-called *small world phenomenon*, with two distinguishing traits —*small distance* between any pair of nodes, and the *clustering effect* that two nodes are more likely to be adjacent if they share a neighbor.

There have been various approaches to model networks that have the small world phenomenon. Progress has been made in analyzing the aspect of small distances by using generalized random graph theory and properties of the power law distribution. However, the clustering effect seems much harder to model.

In 1999, several research groups independently observed that numerous networks such as Internet graphs, call graphs and social networks, etc. all have a *power law* distribution [1,2,5–7,11,13,17,21,23,25,27,33,36]. Namely, the number of nodes of degree k is proportional to $k^{-\beta}$ for some positive exponent β. By using a random graph model for a given degree distribution, it can be shown [16] in a rigorous way that, for example, a random power law graph with exponent β, where $2 < \beta < 3$, almost surely has average distance of order $\log \log n$ and has diameter of order $\log n$. (Note that the average distance is the average of distances between pairs of nodes that are connected and the diameter is the maximum distance between such pairs of nodes.)

To model the clustering effect, most common approaches just add random edges to grid graphs or the like (see Watts and Strogatz [34,35]). Kleinberg [24] introduced the network model of a grid graph with additional random edges joining two nodes u, v with probability proportional to $[d(u, v)]^{-r}$ (where $d(.,.)$ represents the distance in the grid graph and r is a constant that determines the effectiveness of decentralized algorithms for the network). In Kleinberg's model and the model of Watts and Strogatz, the graphs have the same expected degree at every node and do not have a power law degree distribution. Fabrikant, Koutsoupias and Paradimitriou[20] proposed a model of having vertices in the Euclidean plane and adding edges by optimizing the trade-off between (Euclidean) distances and "centrality" in the network. Such grid-based models are quite restrictive and far from satisfactory for modeling webgraphs or biological networks, for example.

The difficulty in reconciling these two aspects – small distance and clustering effect– ressembles philosophically the challenge in physical world concerning the "weak force" and the "strong force". There is no unified model embracing both the weak and strong forces in spite of intense efforts of many great scientists. Random power law graphs are good for modeling the aspect of small distance, but fail miserably for modeling the clustering effect. As a matter of fact, the related graph-theoretical parameters involving small distances and clustering seem to be of an entirely different scale. For example, the clustering effect is quite sensitive to average degree but this is not so for the small (average) distance. Examples of real graphs tell the same story. According to Henzinger [22] at Google, random graphs are good for modeling interdomain hyperlinks but not the local links.

In this paper, we consider a general hybrid graph model that has both aspects of the small world phenomenon. Roughly speaking, a hybrid graph is a union of a global graph (consisting of "long edges" providing small distances) and a local graph (consisting of "short edges" respecting local connections). (Detailed definitions will be given in Section 2.) By using several tools for dealing with random graphs with given expected degree sequences, we will prove that our hybrid graphs have the following properties:

1. Power law degree distribution for a given power β.
2. Small average distance at the same order as that of random graphs.

3. Small diameter at the same order as that of random graphs.
4. Locally highly connected.

We will show that the average distance/diameter is bounded above by $c \log n$ where c depends on the "second-order" average degree (which will be defined later). Consequently, this implies a polylog upper bound for analyzing many distributed algorithms if the network can be well approximated by this general family of hybrid power law graphs.

Motivated by the hybrid graph model, we will give a simple decomposition algorithm. For any real network, the decomposition algorithm identifies the local graph and the global graph. We can then use our theorems to deduce properties of the real network if the local graph satisfies local connectivity conditions and the global graph can be approximated by a random power law graph.

This paper is organized as follows. In Section 2, we give basic definitions for power law graphs and random graphs. In Section 3, we consider local graphs and give the decomposition algorithm. In Section 4, we propose the hybrid graph model by combining a local graph and a random power law graph. We also show that the local graph can be extracted from the hybrid graph with an error estimate of lower order. In Section 5, we describe several useful facts about random graphs with given expected degrees. In particular, we summarize some facts concerning the average distance and diameter of random power law graphs. In Section 6, we establish the desired upper bounds for average distance/diameter for hybrid power law graphs. Section 7 includes further discussions and a number of remarks.

2 Preliminaries

Before we consider the hybrid graphs, we will discuss random graphs with given expected degree sequences and power law degree distribution.

2.1 Random graphs with given expected degrees

We consider a general class of random graphs with given expected degree sequence $\mathbf{w} = (w_1, w_2, \ldots, w_n)$. The probability p_{ij} that there is an edge between vertex v_i and vertex v_j is $w_i w_j \rho$ for any index i and index j. Here we choose ρ to be $(\sum w_i)^{-1}$ and we assume that $\max_i w_i^2 < \sum_k w_k$ so that $p_{ij} \leq 1$ for all i and j. It is then easy to check that the vertex v_i has expected degree w_i. We remark that the assumption $\max_i w_i^2 < \sum_k w_k$ implies that the sequence w_i is graphical (in the sense that it satisfies the necessary and sufficient condition for a sequence to be realized by a graph [18]) except that we do not require the w_i's to be integers. We note that this model allows a non-zero probability for self-loops. The expected number of loops is quite small (of lower order) in comparison with the total number of edges. Consequently, loops have little effect on various graph properties such as average distance, clusterness, etc.

We denote a random graph with a given expected degree sequence $\mathbf{w}$ by $G(\mathbf{w})$. For example, the typical random graph $G(n, p)$ (see [19]) on n vertices and edge density p is just a random graph with expected degree sequence $(pn, pn, \ldots, pn)$. The random graph $G(\mathbf{w})$ is different from the random graphs with a prescribed exact degree sequence (which involve dependency and are hard to analyze). For example, in [30,31], Molloy and Reed obtained results on the sizes of connected components for random graphs with prescribed exact degree sequences which are required to satisfy certain "smoothing" conditions. Our model is also different from the evolution models generated by simple growth rules (such as preferential attachment schemes as in [2,6,10,17]).

2.2 Power law degree distribution

If a graph strictly follows the power law, then the average degree as well as its connectivity (i.e., the distribution of connected components) will be completely determined by the exponent of the power law (see [2]). However, for most realistic graphs, the power law holds only for a certain range of degrees, namely, for the degrees which not too small and not too large. We will consider the following model with the consideration that most examples of massive graphs satisfying power law have exponent $\beta > 2$.

Model $M(n, \beta, d, m)$ where
- n is the number of vertices,
- $\beta > 2$ is the power of the power law,
- d is the expected average degree,
- m is the expected maximum degree (or an upper bound for the range of degrees that obey the power law) and $m^2 = o(nd)$.

We assume that the $i - i_0 + 1$-th vertex v_i has expected degree

$$w_i = ci^{-\frac{1}{\beta-1}}$$

for $i_0 \leq i < n + i_0$. Here c depends on the average degree d and i_0 depends on the maximum expected degree m. It is easy to compute that the number of vertices of expected degree between k and $k + 1$ is of order $c'k^{-\beta}$ where

$c' = c^{\beta-1}(\beta-1)$ as required by the power law. To determine c, we consider

$$\begin{aligned}\mathrm{Vol}(G) = \sum_i w_i &= \sum_{i=i_0}^{n} c i^{-\frac{1}{\beta-1}} \\ &\approx c\frac{\beta-1}{\beta-2} n^{1-\frac{1}{\beta-1}}\end{aligned}$$

Here we assume $\beta > 2$. Since $nd \approx \mathrm{Vol}(G)$, we have

$$c = \frac{\beta-2}{\beta-1} d n^{\frac{1}{\beta-1}} \tag{1}$$

$$i_0 = n\left(\frac{d(\beta-2)}{m(\beta-1)}\right)^{\beta-1} \tag{2}$$

Here equation (2) is deduced from the (cut-off) condition $w_{i_0} = m$ and equation (1).

Let $f(x) = \frac{\beta-2}{\beta-1} d x^{-\frac{1}{\beta-1}}$. The expected degrees (or weights) are just $f(\frac{i}{n})$, $i_0 \le i \le n$.

We will also consider an alternative model $M'(n, \beta, d, m)$, in which each vertex x is assigned a weight $f(y)$, where y is a real number chosen uniformly in the range of (i_0, n). It can be easily shown that two models are equivalent (for $i_0 \ll n$) in the sense that a property that holds for a random graph in M almost surely must hold for M' and vice versa.

2.3 The volume and the second-order average degree

For a subset S of vertices, the k-th volume of S, denoted by $\mathrm{Vol}_k(S)$, is the sum of the k-th power of weights of vertices in S.

$$\mathrm{Vol}_k(S) = \sum_{v_i \in S} w_i^k$$

The expected average degree is defined to be

$$\frac{\mathrm{Vol}_1(G)}{\mathrm{Vol}_0(G)} = \frac{1}{n}\sum_i w_i = d(1+o(1)).$$

We write $\mathrm{Vol}(G) = \mathrm{Vol}_1(G)$. Of particular interest is the second-order average degree $\tilde{d}$ defined by

$$\tilde{d} = \frac{\mathrm{Vol}_2(G)}{\mathrm{Vol}_1(G)} = \frac{\sum_i w_i^2}{\sum_i w_i}.$$

For power law graphs with exponent β, we have [16]

$$\tilde{d} = \frac{\mathrm{Vol}_2(G)}{\mathrm{Vol}_1(G)}$$
$$= \begin{cases} d\frac{(\beta-2)^2}{(\beta-1)(\beta-3)}(1+o(1)) & \text{if } \beta > 3. \\ \frac{1}{4}d\ln m(1+o(1)). & \text{if } \beta = 3. \\ d\frac{(\beta-2)^2}{(\beta-1)(3-\beta)}m^{3-\beta}(1+o(1)) & \text{if } 2 < \beta < 3. \end{cases}$$

3 Local graphs

Roughly speaking, a local graph is locally highly connected. To be precise, here we use two parameters to describe the local connectivity. For any fixed two integers $k \geq 2$ and $l \geq 2$, a graph L is called "locally (k,l)-connected" if for any edge uv, there are at least l edge-disjoint paths (i.e, no two paths share a common edge) with length at most k joining from u to v (including the edge uv). For example, the grid graph $C_n \Box C_n$ is locally $(3,3)$-connected as well as locally $(5,9)$-connected.

By the definition, the union of two locally (k,l)-connected graphs is locally (k,l)-connected. The maximum locally (k,l)-connected subgraph H is the union of all locally (k,l)-connected subgraphs of G. Thus, for any graph G, the maximum locally (k,l)-connected subgraph is unique. We remark that a (k,l)-connected graph is not necessarily connected. For example, the disjoint union of two (k,l)-connected graphs is still (k,l)-connected.

Here is a simple greedy algorithm for finding the maximum locally (k,l)-connected subgraph.

Algorithm(k,l)**:**
For each edge $e = uv$, check whether there are l edge-disjoint paths with length at most k connecting u and v in the current graph G. If not, delete the edge e from G. Then iterate the procedure until no edge can be removed.

Theorem 1. *For any graph G, Algorithm(k,l) finds the unique maximum locally (k,l)-connected subgraph regardless of the order of edges chosen.*

Proof: Let H' be a graph produced by the Algorithm(k,l) where the order that edges are removed is arbitrary. It is sufficient to show $H = H'$.

Let $G = H_0 \supset H_1 \supset H_2 \supset \cdots \supset H_r = H'$ be the sequence of the intermediate subgraphs produced by Algorithm(k,l). We will prove $H \subset H_i$ for all $i = 0, 1, 2, \ldots, r$ by induction on i. It is trivial for $i = 0$ since $H \subset$

$H_0 = G$. Now we assume $H \subset H_i$. For $i+1$, let $e_{i+1} = uv$ be the edge being removed at the $(i+1)$-st stage. It is sufficient to show uv is not an edge of H. Otherwise, there are l edge-disjoint paths of H joining from u to v. Since $H \subset H_i$, these paths are also paths of H_i. According to the algorithm, it can not be removed, which is a contradiction. Thus, we have $H \subset H_{i+1}$ and $H \subset H_r = H'$.

In the other direction, since H' is locally (k, l)-connected, we have $H' \subset H$. H is the maximum subgraph with this property. The proof is complete. □

The edges removed are considered the "global" edges. For certain classes of graphs, the local graph can be almost perfectly recovered as shown by the Figure 1 and 2.

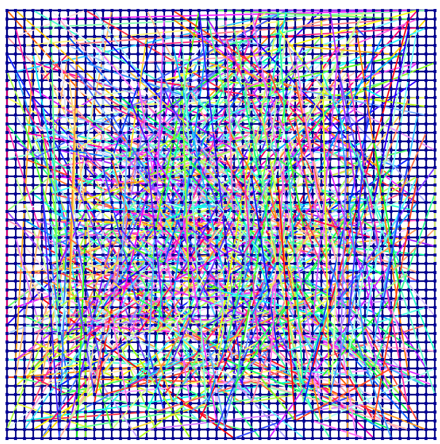

Fig. 1. *A hybrid graph, which contains the grid graph $C_{50}\square C_{50}$ as the local graph, and 528 additional random edges.*

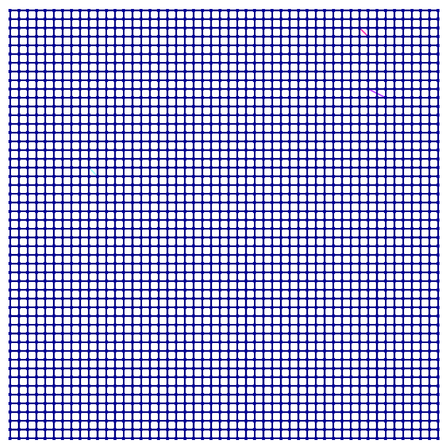

Fig. 2. *After removing all global edges (with $k = l = 3$), the local graph is almost perfectly recovered.*

Another example is the Collaboration Graph of the second kind with 237,426 vertices (as authors of *Math Review*) and about 226,194 edges (each of which is associated with a paper with *exactly* two coauthors), (see http://www.oakland.edu/∼grossman/erdoshp.html for detailed explanations). The local graph L (with $k = l = 3$) has 1979 vertices and 4221 edges. Their degree distributions are showed by Figure 3, 4 and 5.

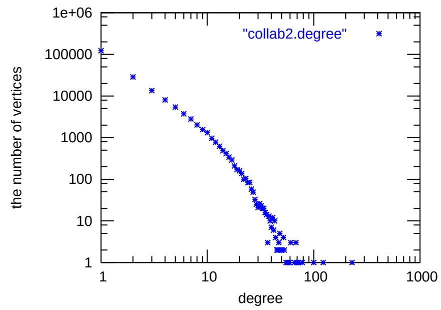

Fig. 3. *The degree distribution of the Collaboration Graph of the second kind.*

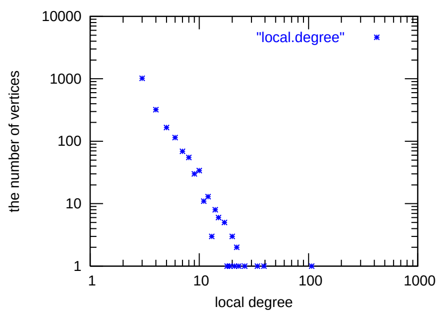

Fig. 4. *The degree distribution of the local graph of the Collaboration Graph of the second kind (with $k = l = 3$) .*

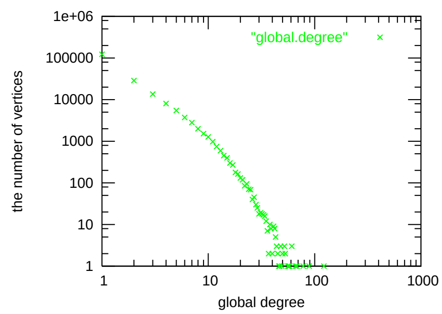

Fig. 5. *The degree distribution of the global graph of the Collaboration Graph of the second kind.*

4 The hybrid power law model

A hybrid graph consists of two parts – a global graph and a local graph. The edge set of the hybrid graph is a disjoint union of the edge set of the global graph and that of the local graph. The related parameters include:

β, the power law exponent,
d, the average degree,
m, the expected maximum degree (or an upper bound for the range of degrees that obey the power law), and
L, the local graph.

We remark that for a given network, all these parameters are straightforward to compute and estimate. Therefore it is quite easy to build a simulation for a network with given parameters.

The hybrid graph $H(n, \beta, d, m, L)$**:**
The local graph L is a locally (k, l)-connected graph with bounded degrees.

The vertex v_i of H has weight w_i where $w_1, w_2, \ldots, w_n$ satisfy a power law distribution with power $\beta > 2$ using the model $M'(n, \beta, d, m)$ in the Section 2.2. We also assume that $d \geq 1$.

For any two points u and v, the probability of having an edge between u and v is denoted by $p(u, v)$, defined as follows:

$$p(u,v) = \begin{cases} 1 & \text{if } uv \text{ is an edge of } L; \\ w_u w_v \rho & \text{otherwise.} \end{cases}$$

We will see that the local graph is quite robust in the sense that it can be almost completely recovered from the hybrid graph.

Theorem 2. *For any fixed constants M, $k \geq 2$, and $l \geq 3$, suppose L is a locally (k,l)-connected graph with degrees bounded by M. Let L' be the maximum locally (k,l)-connected subgraph in the hybrid graph $H(n, \beta, d, m, L)$ with $m = o(n^{\frac{1}{2}-\frac{1}{l}})$. Then L' satisfies*

1. *$L \subset L'$. The expected number of edges in $L' \backslash L$ is small, i.e., $e(L') - e(L) = O(m) = o(\sqrt{n})$.*
2. *The degree of L' is almost surely bounded above by $M + \lceil \frac{l}{2} \rceil - 1$.*
3. *The diameter $D(L')$ of L' is almost surely $(1 + o(1))D(L)$ if the diameter $D(L)$ is sufficiently large.*

Proof: From the definitions, we have $L \subset L'$. Thus, $D(L') \leq D(L)$. In the other direction, we consider edges in L' but not in L, which we call the *surviving* edges. We call the distance of two vertices in L the *local distance*, denoted by d_L. The neighborhood of a vertex in L is said to be a *local neighborhood*. A i-th local neighborhood of v consists of all vertices within local distance i from v. We will prove the following:

Claim: Almost surely all surviving edges uv have endpoints with local distance $d_L(u, v)$ at most k.

For any vertices u and v, if $d_L(u, v) > k$, any path of length at most k in L' from u to v must contain at least one surviving edge. Since this edge uv survives after the algorithm terminates, there exist at least l edge-disjoint edges in L' from the i-th local neighborhood of u to the j-th local neighborhood of v with some $i + j = k - 1$. Since the local degrees are bounded by M, the number of vertices in the i-th local neighborhood of u is at most

$$\sum_{s=0}^{i} M^s = \frac{M^{i+1}}{M-1} \leq 2M^i.$$

Similarly, the number of vertices in the j-th local neighborhood of u is at most $2M^j$. There are at most $2M^i \times 2M^j = 4M^{k-1}$ pairs of such vertices. For each pair, the probability of being randomly chosen for the hybrid graph is less than $m^2\rho$. Thus, the probability that uv survives is at most

$$\binom{4M^{k-1}}{l}(m^2\rho)^l = o(\frac{1}{n^2}).$$

Therefore, almost surely, all surviving edges have endpoints with local distance at most k.

Now we consider the expected number of surviving edges, which almost surely have endpoints within local distance at most k. We choose a vertex u. There are at most $2M^k$ vertices with local distance at most k from u. The expected number of surviving edge uv with $d_L(u,v) \leq k$ is at most $\sum_u 2M^k w_u m\rho = 2M^k m$. For m large, the number of surviving random edges is well-concentrated on its expected value.

For any fixed u, we examine the number of surviving edges uv which are incident to u. Since almost surely v is within local distance k from u, there are at most $2M^k$ of possible v's. The probability that there are at least $\lceil \frac{l}{2} \rceil$ v's with uv surviving is at most

$$\binom{2M^k}{\lceil \frac{l}{2} \rceil}(m^2\rho)^{\lceil \frac{l}{2} \rceil} = o(\frac{1}{n}).$$

Thus, almost surely the degree of L' is at most $M + \lceil \frac{l}{2} \rceil - 1$.

Let $g(n)$ be a (very) slowly growing function of n, which approaches infinity as n increases. There are at most $4M^{2g(n)}$ pairs of vertices within local distance $g(n)$ from any given vertex u. The probability that l surviving edges are within local distance $g(n)$ from any given vertex u is at most

$$\binom{4M^{2g(n)}}{l}(m^2\rho)^l = o(\frac{1}{n^2}),$$

for some slowly growing function $g(n) = o(\log \frac{n^{\frac{1}{2}-\frac{1}{l}}}{m})$. Almost surely, for all vertex u, there are at most $l-1$ surviving edges with local distance at most k from u.

Let (u,v) be a pair of vertices with $d_L(u,v) = D(L)$. The distance between u and v in the hybrid graph can be reduced (from the local distance of u and v) by surviving edges. Each surviving edge can reduce the distance from u to v by at most $k-1$. The total number of surviving edges which can be used on the path from u to v is at most $\frac{l}{g(n)} = o(1)$. Hence $d_{L'}(u,v) \geq (1 - \frac{kl}{g(n)})d_L(u,v) = (1-o(1))D(L)$. Thus, the diameter $D(L')$ is at least $(1-o(1))D(L)$. This completes the proof of Theorem 2 □

5 Several facts concerning random power law graphs

In this section we state several useful facts for random power law graphs $G(\mathbf{w})$ with given expected degree sequence $\mathbf{w}$. Proofs of these facts can be found in [16].

The expected degree sequence $\mathbf{w}$ for a graph G on n vertices in $G(\mathbf{w})$ is said to be *strongly sparse* if we have the following :
(i) The second order average degree $\tilde{d}$ satisfies $0 < \log \tilde{d} \ll \log n$.
(ii) For some constant $c > 0$, all but $o(n)$ vertices have expected degree w_i satisfying $w_i \geq c$. The average expected degree $d = \sum_i w_i/n$ is strictly greater than 1, i.e., $d > 1 + \epsilon$ for some positive value ϵ independent of n.

The expected degree sequence $\mathbf{w}$ for a graph G on n vertices in $G(\mathbf{w})$ is said to be *admissible* if the following condition holds, in addition to the assumption that $\mathbf{w}$ is strongly sparse.
(iii) There is a subset U satisfying:

$$\mathrm{Vol}_2(U) = (1 + o(1))\mathrm{Vol}_2(G) \gg \frac{\mathrm{Vol}_3(U) \log \tilde{d} \log \log n}{\tilde{d} \log n}.$$

The expected degree sequence $\mathbf{w}$ for a graph G on n vertices is said to be *specially admissible* if (i) is replaced by (i') and (iii) is replaced by (iii'):
(i') $\log \tilde{d} = O(\log d)$.
(iii') There is a subset U satisfying $\mathrm{Vol}_3(U) = O(\mathrm{Vol}_2(G))\frac{\tilde{d}}{\log \tilde{d}}$, and $\mathrm{Vol}_2(U) > d\mathrm{Vol}_2(G)/\tilde{d}$.

Fact 1 *For a random graph G with admissible expected degree sequence $(w_1, \ldots, w_n)$, the average distance is almost surely $(1 + o(1))\frac{\log n}{\log \tilde{d}}$.*

Fact 2 *For a random graph G with a specially admissible degree sequence $(w_1, \ldots, w_n)$, the diameter is almost surely $\Theta(\log n/ \log \tilde{d})$.*

Fact 3 *For a power law random graph with exponent $\beta > 3$ and average degree d strictly greater than 1, almost surely the average distance is $(1 + o(1))\frac{\log n}{\log \tilde{d}}$ and the diameter is $\Theta(\log n)$.*

Fact 4 *Suppose a power law random graph with exponent β has average degree d strictly greater than 1 and maximum degree m satisfying $\log m \gg \log n/ \log \log n$. If $2 < \beta < 3$, almost surely the diameter is $\Theta(\log n)$ and the average distance is at most $(2 + o(1))\frac{\log \log n}{\log(1/(\beta-2))}$.*

For the case of $\beta = 3$, the power law random graph has diameter almost surely $\Theta(\log n)$ and has average distance $\Theta(\log n / \log\log n)$.

The proofs of the above facts use the following lemmas concerning the distances and neighborhood expansions in $G(\mathbf{w})$. These lemmas (as proved in [15]) are useful later for proving the main theorems in the next section.

Lemma 1. *In a random graph G in $G(\mathbf{w})$ with a given expected degree sequence $\mathbf{w} = (w_1, \ldots, w_n)$, for any fixed pairs of vertices (u, v), the distance $d(u, v)$ between u and v is greater than $\left\lfloor \frac{\log \mathrm{Vol}(G) - c}{\log \tilde{d}} \right\rfloor$ with probability at least $1 - \frac{w_u w_v}{\tilde{d}(\tilde{d}-1)} e^{-c}$.*

Lemma 2. *In a random graph $G \in G(\mathbf{w})$, for any two subsets S and T of vertices, we have*

$$\mathrm{Vol}(\Gamma(S) \cap T) \geq (1 - 2\epsilon)\mathrm{Vol}(S)\frac{\mathrm{Vol}_2(T)}{\mathrm{Vol}(G)}$$

with probability at least $1 - e^{-c}$ where $\Gamma(S) = \{v \ : \ v \sim u \in S \text{ and } v \notin S\}$, provided $\mathrm{Vol}(S)$ *satisfies*

$$\frac{2c\mathrm{Vol}_3(T)\mathrm{Vol}(G)}{\epsilon^2 \mathrm{Vol}_2^2(T)} \leq \mathrm{Vol}(S) \leq \frac{\epsilon \mathrm{Vol}_2(T)\mathrm{Vol}(G)}{\mathrm{Vol}_3(T)} \tag{3}$$

Lemma 3. *For any two disjoint subsets S and T with* $\mathrm{Vol}(S)\mathrm{Vol}(T) > c\mathrm{Vol}(G)$, *we have*

$$Pr(d(S, T) > 1) < e^{-c}$$

where $d(S, T)$ denotes the distance between S and T.

6 The diameter of the hybrid model

Most local graphs have large diameters and large average distances. For example, the average distance of the grid graph on n vertices is $O(\sqrt{\frac{n}{\log n}})$. However, with additional "hyperlinks", (e.g., edges from the global random power law graph), the average distance of the hybrid graph can be significantly reduced.

In a hybrid graph H, let G denote its global power law graph as defined in Section 4. Let $\mathbf{w} = (w_1, w_2, \ldots, w_m)$ denote the degree sequence of G. We will say that vertex v_i has weight w_i and we recall that for a subset S of vertices, we have $\mathrm{Vol}(S) = \sum_{v_i \in S} w_i$ and $\mathrm{Vol}(G) = \sum w_i$. Also for $k \geq 1$, we have $\mathrm{Vol}_k(S) = \sum_{v_i \in S} w_i^k$. In particular, the second order average degree $\tilde{d}$ is just $\mathrm{Vol}_2(G)/\mathrm{Vol}(G)$. The following are immediate consequences of Fact 3 and Fact 4.

Theorem 3. *For a hybrid graph $H(n, \beta, d, m, L)$ with $\beta > 3$, almost surely, the average distance is $(1 + o(1))\frac{\log n}{\log \tilde{d}}$ and the diameter is $O(\log n)$.*

Theorem 4. *For a hybrid graph $H(n, \beta, d, m, L)$ with $2 < \beta < 3$, almost surely, the average distance is $O(\log \log n)$ and the diameter is $O(\log n)$.*

For a hybrid graph $H(n, \beta, d, m, L)$ with $\beta = 3$, almost surely, the average distance is $O(\log n / \log \log n)$ and the diameter is $O(\log n)$.

For the range of $2 < \beta < 3$, the power law graphs include many real networks. We can further reduce the diameter if additional conditions are satisfied. A local graph L is said to have isoperimetric dimension δ if for every vertex v in L and every integer $k < (\log \log n)^{1/\delta}$, there are at least k^δ vertices in L of distance k from v. For example, the grid graph in the plane has isoperimetric dimension 2. The d-dimensional grid graph has isoperimetric dimension d.

Theorem 5. *In a hybrid graph $H(n, \beta, d, m, L)$ with $2 < \beta < 3$, suppose that the local graph has isoperimetric dimension δ with $\delta \geq \log \log n/(\log \log \log n)$. Then almost surely, the diameter is $O(\log \log n)$.*

The main idea of the proof of Theorem 5 is to use the "octopus" structure of the random powerlaw graph with exponent β between 2 and 3. The proof is quite similar to that in [16] except that here we have the additional help from the local graph. For the sake of completeness, we include the proof here.

Proof of Theorem 5:

First, we define the core of a power law graph with exponent β to be the set S_t of vertices of degree at least $t = n^{1/\log \log n}$.

Claim 1: The diameter of the core is almost surely $O(\log \log n)$. This follows from the fact that the core contains an Erdős-Renyi graph $G(n', p)$ with $n' = cnt^{1-\beta}$ and $p = t^2/\mathrm{Vol}(G)$. From [19], this subgraph is almost surely connected. Using a result in [14], the diameter of this subgraph is at most $\frac{\log n'}{\log pn'} = (1 + o(1))\frac{\log n}{(3-\beta)\log t} = O(\log \log n)$.

Claim 2: Almost all vertices with degree at least $\log n$ are almost surely within distance $O(\log \log n)$ from the core. To see this, we start with a vertex u_0 with degree $k_0 \geq \log^C n$ for some constant $C = \frac{1.1}{(\beta-2)(3-\beta)}$. By applying Lemma 3, with probability at least $1-n^{-3}$, u_0 is a neighbor of some u_1 with degree $k_1 \geq (k_0/\log^C n)^{1/(\beta-2)^s}$. We then repeat this process to find a path with vertices $u_0, u_1, \ldots, u_s$, and the degree k_s of u_s satisfies $k_s \geq (k_0/\log^C n)^{1/(\beta-2)^s}$ with probability $1-n^{-2}$. By choosing s to satisfy $\log k_s \geq \log n/\log\log n$, we are done.

Claim 3: Each vertex v is within distance $O(\log\log n)$ from a vertex of degree at least $\log^C n$.

Proof of Claim 3: The main tools are Lemma 2. Let S be i-th neighborhood of u, consisting of all vertices within distance i_0 from u where $i_0 = \log\log n$. Let $T = S(w_{\min}, a)$ denote the set of vertices with weights between $w_{\min}$ and $aw_{\min}$. Here a is some large value to be chosen later. We have

$$\begin{aligned}\mathrm{Vol}(T) &\approx nd(1-a^{2-\beta}).\\ \mathrm{Vol}_2(T) &\approx nd^2(1-\frac{1}{\beta-1})^2\frac{\beta-1}{3-\beta}a^{3-\beta}\\ \mathrm{Vol}_3(T) &\approx nd^3(1-\frac{1}{\beta-1})^3\frac{\beta-1}{4-\beta}a^{4-\beta}\end{aligned}$$

To apply Lemma 2, $\mathrm{Vol}(\Gamma(S))$ must satisfy:

$$\begin{aligned}\mathrm{Vol}(\Gamma(S)) &\geq \frac{2c}{\epsilon^2}\frac{\mathrm{Vol}_3(T)}{\mathrm{Vol}_2^2(T)}\mathrm{Vol}(G)\\ &\approx \frac{2c}{\epsilon^2}\frac{(3-\beta)^2}{(\beta-2)(4-\beta)}a^{\beta-2}\end{aligned}$$

and

$$\begin{aligned}\mathrm{Vol}(\Gamma(S)) &\leq \epsilon\frac{\mathrm{Vol}_2(T)}{\mathrm{Vol}_3(T)}\mathrm{Vol}(G)\\ &\approx \epsilon\frac{(\beta-2)(3-\beta)}{(\beta-1)(4-\beta)a}n.\end{aligned}$$

Both the above equations are easy to satisfy by using the assumption on the local graph. Namely, we can select $a = 10$, $c = 3\log n$ and for each vertex u,

$$\mathrm{Vol}(\Gamma_{i_0}(u)) \geq \frac{60}{\epsilon^2(\beta-2)}\log n.$$

By Lemma 2, with probability at least $1-e^{-c} = 1-\frac{1}{n^3}$, the volume of $\Gamma_i(u)$ for $i > i_0$ will grow at a rate greater than

$$(1-2\epsilon)\frac{\mathrm{Vol}_2(T)}{\mathrm{Vol}(G)} \approx \frac{(1-2\epsilon)d(\beta-2)^2}{2(\beta-1)(3-\beta)}a^{3-\beta},$$

if $\Gamma_i(u)$ has volume not too large ($< \sqrt{n}$). After at most $(1+o(1))\frac{2\,\mathrm{loglog}n}{(3-\beta)\log a} = O(\log\log n)$ steps, the volume of the reachable vertices is at least $\log^2 n$. Lemma 3 then implies that with one additional step we can reach a vertex of weight $\log^C n$ with probability at least $1 - e^{-\log^2 n}$ The total number of steps is at most

$$i_0 + O(\log\log n) + 1 = O(\mathrm{loglog}n).$$

The total failure probability for u to reach a vertex of weight at least $\log^C n$ is at most

$$O(\log\log n)\frac{1}{n^3} + e^{-O(\log^2 n)} = o(\frac{1}{n^2}).$$

Thus, the total failure probability that some vertex u can not reach a vertex of weight at least $\log^C n$ is at most

$$o(1) + O(\log\log n)O(\frac{1}{n}) + ne^{-O(\log^2 n)} = o(1).$$

Claim 3 is proved. □

This completes the proof of Theorem 5. □

The proof of the above theorem implicitly implies the following results:

Theorem 6. *In a hybrid graph $H(n, \beta, d, m, L)$ with $2 < \beta < 3$, suppose that the local graph has isoperimetric dimension δ. Then almost surely, the diameter is $O((\log n)^{1/\delta})$.*

Theorem 7. *In a hybrid graph $H(n, \beta, d, m, L)$ with $2 < \beta < 3$, suppose that every vertex is within distance $\log\log n$ of some vertex of degree $\log n$. Then almost surely, the diameter is $O(\log\log n)$.*

7 Concluding remarks

In this paper, we consider the hybrid model for further understanding the "landscape" of real networks. Here we mention a number of remarks concerning the flexibility and possible extensions of our model.

1. In our hybrid model, the global graph was chosen to be a random graph with given degree distribution satisfying a power law. If the global graph is to be taken to be an admissible graph or specially admissible graph (as defined in Section 5), similar results on average distance and diameter can be established by using methods in the the proofs of Theorem 5. There are several reasons for selecting the global graph to be a power law

graph. Namely, many real networks have power law degree distribution. In addition, random power law graph G has the "scale-free" property [3] in the sense that if a fraction of vertices or edges are deleted from G, the remaining graph is still a power law graph with the same exponent (but with different average degree).

2. It is of interest to further analyze the local graphs for various classes of networks. In addition to local connectivity, are there other distinct properties that local graphs have? One such example is the isoperimetric dimension (as defined in Section 6) or its variations. Different types of networks (Internet graphs versus biological graphs, and so on) can have different kinds of local graphs. Are there good characterizations for different local graphs? In particular, are there special characterizations for local graphs for networks arising in epidemics and percolation?

3. The local graph in our hybrid model is a (k, l)-connected graph, with parameters k and l which can be chosen to suit the actual network under consideration. We note that (k, l)-connected graphs include the grid graphs, disjoint union of grid graphs, and grid graphs of higher dimensions, depending on the choice of k and l. In fact, by appropriately choosing several pairs of k and l, the algorithm given in Section 3 can result in a (k, l)-connected subgraph with a number of distinct connected components and thereby identify local "communities" within the (large) network. For example, for $k = l = 3$, the local graph of Collaboration graph of the second kind is the disjoint union of 149 (non-trivial) components as shown in Figure 6. Each component can be viewed as a community.

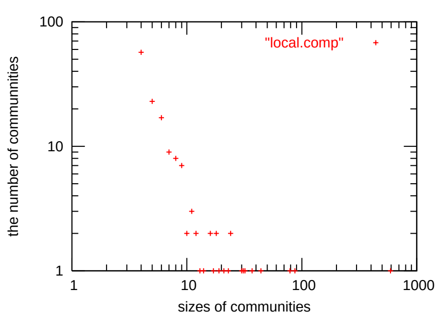

Fig. 6. *The distribution of the sizes of local communities in the local graph (with $k = l = 3$) of the Collaboration Graph of the second kind.*

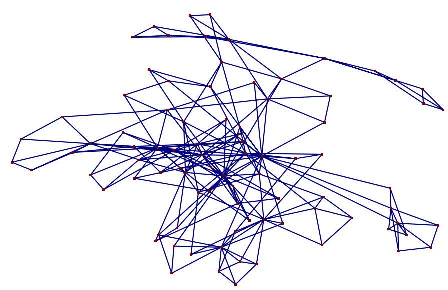

Fig. 7. *A "community" of size 87 in the Collaboration Graph of the second kind.*

8 Acknowledgments

This research was supported in part by NSF Grants DMS 0100472 and ITR 0205061.

References

1. L. A. Adamic and B. A. Huberman, Growth dynamics of the World Wide Web, *Nature*, **401**, September 9, 1999, pp. 131.
2. W. Aiello, F. Chung and L. Lu, A random graph model for massive graphs, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, (2000) 171-180.
3. W. Aiello, F. Chung and L. Lu, Random evolution in massive graphs, Extended abstract appeared in *The 42th Annual Symposium on Foundation of Computer Sciences*, October, 2001. Paper version appeared in *Handbook on Massive Data Sets*, (Eds. J. Abello, et. al.), Kluwer Academic Publishers (2002), 97-122.
4. N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley and Sons, New York, 1992.
5. R. B. R. Azevedo and A. M. Leroi, A power law for cells, *Proc. Natl. Acad. Sci. USA*, vol. **98**, no. 10, (2001), 5699-5704.
6. A.-L. Barabási and R. Albert, Emergence of scaling in random networks, Science **286** (1999) 509-512.
7. A. Barabási, R. Albert, and H. Jeong, Scale-free characteristics of random networks: the topology of the world wide web, *Physica* A 272 (1999), 173-187.
8. E. A. Bender and E. R. Canfield, The asymptotic number of labelled graphs with given degree sequences, *J. Combinat. Theory (A)*, **24**, (1978), 296-307.
9. B. Bollobás, *Random Graphs*, Academic, New York, 1985.
10. B. Bollobás, O. Riordan, J. Spencer and G. Tusnády, The Degree Sequence of a Scale-Free Random Graph Process, Random Structures and Algorithms, Vol. **18**, no. 3 (2001), 279-290.
11. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tompkins, and J. Wiener, "Graph Structure in the Web," *proceedings of the WWW9 Conference*, May, 2000, Amsterdam. Paper version appeared in *Computer Networks* **33**, (1-6), (2000), 309-321.
12. CAIDA, the Cooperative Association for Internet Data Analysis, http://www.caida.org/
13. K. Calvert, M. Doar, and E. Zegura, Modeling Internet topology. *IEEE Communications Magazine*, **35(6)** (1997) 160-163.
14. F. Chung and L. Lu, The diameter of random sparse graphs, *Advances in Applied Math.*, **26** (2001), 257-279.
15. F. Chung and L. Lu, Connected components in a random graph with given degree sequences, *Annals of Combinatorics*, **6** (2002), 125-145.
16. F. Chung and L. Lu, Average distances in random graphs with given expected degree sequences, *Proceedings of National Academy of Science*, **99** (2002), 15879-15882.
17. C. Cooper and A. Frieze, On a general model of web graphs, *Random Structures and Algorithms* Vol. **22**, (2003), 311-335.

18. P. Erdős and T. Gallai, Gráfok előírt fokú pontokkal (Graphs with points of prescribed degrees, in Hungarian), *Mat. Lapok* **11** (1961), 264-274.
19. P. Erdős and A. Rényi, On random graphs. I, *Publ. Math. Debrecen* **6** (1959), 290-291.
20. A. Fabrikant, E. Koutsoupias and C. H. Papadimitriou, Heuristically optimized trade-offs: a new paradigm for power laws in the Internet, *Automata, Languages and Programming*, 110-122, *Lecture Notes in comput. Sci., 2380*, Springer, Berlin, 2002.
21. M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationships of the Internet topology, *Proceedings of the ACM SIGCOM Conference*, Cambridge, MA, 1999.
22. M. Henzinger, private communication.
23. S. Jain and S. Krishna, A model for the emergence of cooperation, interdependence, and structure in evolving networks, *Proc. Natl. Acad. Sci. USA*, vol. **98**, no. 2, (2001), 543-547.
24. J. Kleinberg, The small-world phenomenon: An algorithmic perspective, *Proc. 32nd ACM Symposium on Theory of Computing*, 2000.
25. J. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, The web as a graph: Measurements, models and methods, *Proceedings of the International Conference on Combinatorics and Computing*, 1999.
26. S. R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, Trawling the web for emerging cyber communities, *Proceedings of the 8th World Wide Web Conference*, Toronto, 1999.
27. S. R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, Extracting large-scale knowledge bases from the web, *Proceedings of the 25th VLDB Conference*, Edinburgh, Scotland, 1999.
28. L. Lu, The diameter of random massive graphs, *Proceedings of the Twelfth ACM-SIAM Symposium on Discrete Algorithms*, (2001) 912-921.
29. C. McDiarmid, Concentration. Probabilistic methods for algorithmic discrete mathematics, *Algorithms Combin.*, **16**, Springer, Berlin, (1998) 195–248.
30. M. Molloy and B. Reed, A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, Vol. **6**, no. 2 and 3 (1995), 161-179.
31. M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence, *Combin. Probab. Comput.* **7**, no. 3 (1998), 295-305.
32. S. Milgram, The small world problem, *Psychology Today*, **2** (1967), 60-67.
33. M. E. J., Newman, The structure of scientific collaboration networks, *Proc. Natl. Acad. Sci. USA*, vol. **98**, no. 2, (2001), 404-409.
34. D. J. Watts, *Small Worlds — The Dynamics of Networks between Order and Randomness*, Princeton University Press, New Jersey, 1999.
35. D. J. Watts and S. H. Strogats, Collective dynamics of 'small world' networks, *Nature* **393**, 440-442.
36. E. Zegura, K. Calvert, and M. Donahoo, A quantitative comparison of graph-based models for Internet topology. *IEEE/ACM Transactions on Networking*, **5** (6), (1997), 770-783.