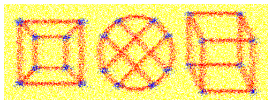
Journal of Graph Algorithms and Applications
http://jgaa.info/ vol. 23, no. 4, pp. 625–651 (2019)
DOI: 10.7155/jgaa.00505

# Density decompositions of networks

*Glencora Borradaile*[1] *Theresa Migler*[2] *Gordon Wilfong*[3]

[1]School of EECS, Oregon State University
[2]CSSE Department, California Polytechnic State University, San Luis Obispo
[3]Nokia Bell Labs

**Abstract**

We introduce a new topological descriptor of a graph called the density decomposition which is a partition of the vertices of a graph into regions of uniform density. The decomposition we define is unique in the sense that a given graph has exactly one density decomposition. The number of vertices in each partition defines a density distribution which we find is measurably similar to the degree distribution of given *real-world* networks (social, internet, etc.) and measurably dissimilar in synthetic networks (preferential attachment, small world, etc.).

We also show how to build networks having given density distributions, which gives us further insight into the structure of real-world networks.

| Submitted: | Reviewed: | Revised: | Accepted: |
|---|---|---|---|
| February 2018 | June 2019 | July 2019 | July 2019 |
| | Final: July 2019 | Published: September 2019 | |
| Article type: Regular paper | | Communicated by: U. Brandes | |

*E-mail addresses:* glencora@eecs.oregonstate.edu (Glencora Borradaile) tmigler@calpoly.edu (Theresa Migler) gordon.wilfong@nokia-bell-labs.com (Gordon Wilfong)

# 1 Introduction

A better understanding of the topological properties of real-world or naturally-occurring networks can be advantageous for two major reasons. First, knowing that a network has certain properties, e.g., bounded degree or planarity, can sometimes allow for the design of more efficient algorithms for extracting information about the network or for the design of more efficient distributed protocols to run on the network. Second, it can lead to methods for generating artificial networks that more accurately match the properties of real-world networks thus allowing for more accurate predictions of future growth of the network and more accurate simulations of distributed protocols running on such a network.

We show that graphs decompose naturally into regions of uniform density, a *density decomposition*. The decomposition we define is unique in the sense that a given graph has exactly one density decomposition. The number of vertices in each region defines a distribution of the vertices according to the density of the region to which they belong, that is, a *density distribution* (Section 2). Although density is closely related to degree, we find that the density distribution of a particular graph is not necessarily similar to the degree distribution of that graph. For example, in many generated networks, such as those generated by popular network models (e.g. preferential attachment and small worlds), the density distribution is very different from the degree distribution (Section 3.1). On the other hand, for *all* of the real-world networks (social, internet, etc.) in our data set, the density and degree distributions are measurably similar (Section 3). Similar conclusions can be drawn using the notion of *k-cores* [33], which we discuss in Section 2.3.

As others before us, we would like a method for generating networks having certain properties that are believed to match those of real-world networks. Having observed that many real-world networks have the property that their degree distributions match their density distributions, we describe a method for generating networks that have that property. We develop an abstract model, that, given a particular density distribution, produces a network having that density distribution (Section 4). Applied naïvely, given a density distribution of a real-world network, this model generates networks with realistic average path lengths (average number of hops between pairs of vertices) and degree distributions; that is similar to the given real-world network (Section 4.1). In addition to having short average path lengths, large-scale, real-world networks also tend to have high *clustering coefficients* [29]. The clustering coefficient of a vertex $v$ is the ratio of the number of pairs of neighbors of $v$ that are connected to the number of pairs of neighbors of $v$; the clustering coefficient of a graph is the average clustering coefficient of its vertices. Our model, naïvely applied, unfortunately, but not surprisingly, results in networks with very low clustering coefficients. However, we show that applying the abstract model in a more sophisticated manner, using ideas from the small world model of Watts and Strogatz [37], results in much higher clustering coefficients (Section 4.2) suggesting that real-world networks may indeed be *hierarchies of small worlds*. Our hierarchies of small worlds specification is just one way to tune our abstract

model; our model is quite flexible, allowing for the easy incorporation of other network generation techniques, which we discuss in Section 4.4.

## 1.1 Related work

In this paper we consider undirected graphs, $G = (V, E)$. The *density* of a graph is the ratio of the number of edges to the number of vertices, $\frac{|E|}{|V|}$.

The *density decomposition* of a graph is a partition of the vertices into regions of uniform density [9]. We can obtain such a decomposition by first orienting the edges of the graph in an *egalitarian*[1] manner. Then we partition the vertices based on their indegree and connectivity in this orientation.

Egalitarian orientations are one example of *fair orientations* that have been studied frequently in the past. For example, Venkateswaran shows how to find an orientation that minimizes the maximum indegree of any vertex [39]. Asahiro, Miyano, Ono, and Zenmyo consider the edge-weighted version of this problem [3]. They give a combinatorial $\{\frac{w_{max}}{w_{min}}, (2-\epsilon)\}$-approximation algorithm where $w_{max}$ and $w_{min}$ are the maximum and minimum weights of edges respectively, and $\epsilon$ depends on the average weights of the input graph [3]. Klostermeyer considers the problem of reorienting edges (rather than whole paths) so as to create graphs with given properties, such as strongly connected graphs and acyclic graphs [19]. De Fraysseix and de Mendez show that they can find an indegree assignment of the vertices given particular properties [13]. Biedl, Chan, Ganjali, Hajiaghayi, and Wood give a $\frac{13}{8}$-approximation algorithm for finding an ordering of the vertices such that for each vertex $v$, the neighbors of $v$ are as evenly distributed to the right and left of $v$ as possible [8]. For the purpose of deadlock prevention [38], Wittorff describes a heuristic for finding an acyclic orientation that minimizes the sum over all vertices of the function $\delta(v)$ choose 2, where $\delta(v)$ is the indegree of vertex $v$ [39].

Often it is of interest to find the *densest subgraph* of a graph, a set of vertices, $S$, such that the density of the induced subgraph on $S$, $\frac{|E(S)|}{|S|}$, is maximized. In our work we show that the density decomposition can isolate the densest subgraph. The densest subgraph problem has been studied a great deal. Goldberg gives an algorithm to find the densest subgraph in polynomial time using network flow techniques [15]. There is a 2-approximation for this problem that runs in linear time [10]. As a consequence of our decomposition, we find a subgraph that has density no less than the density of the densest subgraph minus one.

There are algorithms to find dense subgraphs in the streaming model [4, 14]. There are algorithms that find all densest subgraphs in a graph (there could be many such subgraphs) [32].

[1] An egalitarian orientation is one in which the indegrees of the vertices are as balanced as possible as allowed by the topology of the graph.

## 2 The density decomposition

In order to obtain the density decomposition (a partition of the vertices into regions of uniform density) of a given undirected graph we first orient the edges of this graph in an egalitarian manner. Then we partition the vertices based on their indegree and connectivity in this orientation.

The following procedure, the PATH-REVERSAL algorithm, finds an egalitarian orientation [9]. A *reversible path* is a directed path from a vertex $v$ to a vertex $u$ such that the indegree of $v$, $\delta(v)$, is greater than the indegree of $u$ plus one: $\delta(v) > \delta(u) + 1$

Arbitrarily orient the edges of the graph.
While there is a reversible path
    reverse this path.

Since we are only reversing paths between vertices with differences in indegree of at least 2, this procedure converges; the running time of this algorithm is quadratic [9]. The orientation resulting from this termination condition suggests a hierarchical decomposition of its vertices: Let $k$ be the maximum indegree in the orientation. *Ring* $k$, denoted $R_k$, contains all vertices of indegree $k$ and all vertices that reach vertices of indegree $k$. By the termination condition of the above procedure, only vertices of indegree $k$ or $k-1$ are in $R_k$. Iteratively, given $R_k, R_{k-1}, \ldots,$ and $R_{i+1}$, $R_i$ contains all the remaining vertices with indegree $i$ along with all the remaining vertices that reach vertices with indegree $i$. Vertices in $R_i$ must have indegree $i$ or $i-1$ by the termination condition of the procedure. By this definition, an edge between a vertex in $R_i$ and a vertex in $R_j$ is directed from $R_i$ to $R_j$ when $i > j$ and all the isolated vertices are in $R_0$. See Figure 1 for an example of the density decomposition of a graph.

Recall that we define the *density* of a graph to be the ratio of the number of edges to the number of vertices, $\frac{|E|}{|V|}$. This definition of density is closely related to vertex degree (the number of edges adjacent to a given vertex): the density of a graph is equal to half the average total degree.

We *identify* a set $S$ of vertices in a graph by merging all the vertices in $S$ into a single vertex $s$ and removing any self-loops (corresponding to edges of the graph both of whose endpoints were in $S$). Our partition $R_k, R_{k-1}, \ldots, R_0$ induces regions of uniform density in the following sense:

**Density Property** For any $i = 0, \ldots, k$, identifying the vertices in $\cup_{j>i} R_j$ and deleting the vertices in $\cup_{j<i} R_j$ leaves a graph $G$ whose density is in the range $(i-1, i]$ (for $|R_i|$ sufficiently large).

In particular, $R_k$ isolates a *densest* region in the graph. Consider the graph $G_i$ formed by identifying the vertices $\cup_{j>i} R_j$ and deleting the vertices in $\cup_{j<i} R_j$; this graph has one vertex (resulting from identifying the vertices $\cup_{j>i} R_j$) of indegree 0 and $|R_i|$ vertices of indegree $i$ of $i-1$, at least one of which must

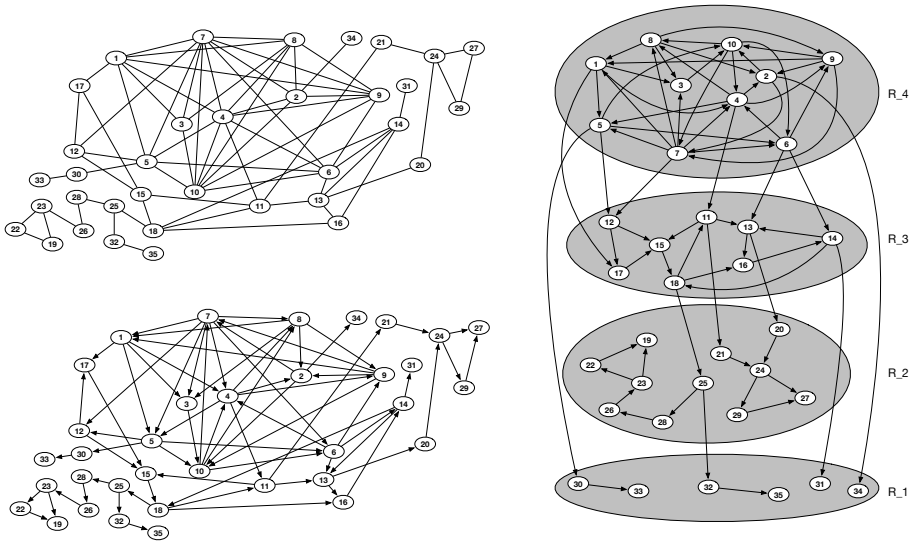


Figure 1: The upper left graph, $G$, is an example input graph. The bottom left graph is the graph $G$ with an egalitarian orientation imposed on its edges. Notice that the maximum indegree in this orientation is 4. Therefore the topmost ring for this graph is $R_4$. The graph on the right illustrates the density decomposition for $G$.

have indegree $i$. Therefore, for any $i$, the density of $G_i$ is at most $i$ and density at least

$$\frac{(|R_i|-1)(i-1)+i}{|R_i|+1} \xrightarrow{|R_i|\gg i} i-1.$$

The relationship between density and this decomposition is much stronger. Let $k$ denote the maximum index of a non-empty ring. In Section 2.1, we show the following properties:

**Property D1** The density decomposition of a graph is unique and does not depend on the starting orientation.

**Property D2** The density of a densest subgraph is at most $k$. That is, there is no denser region $R_j$ for $j > k$.

**Property D3** Every densest subgraph contains only vertices of $R_k$.

These properties allow us to unequivocally describe the density structure of a graph. We summarize the density decomposition by the *density distribution*: $(|R_0|, |R_1|, \ldots |R_{k-1}|, |R_k|)$, i.e. the number of vertices in each region of uniform density. We will refer to a vertex in $R_i$ as having *density rank* $i$.

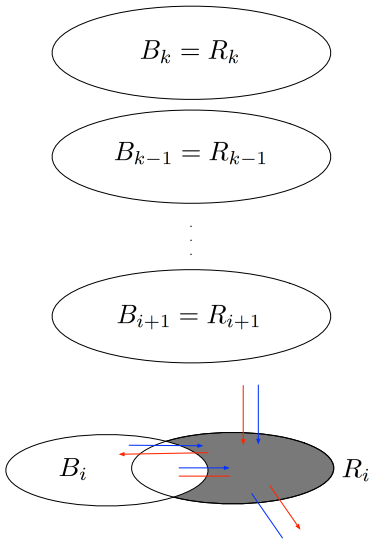


Figure 2: Edges incident to $S$ (grey region) in the proof of Theorem 1 in red and blue orientations. $S$ is the region in gray. Unoriented edges indicate that the orientation could be in either direction.

## 2.1 Density and the Density Decomposition

We now prove that the partition of the rings does not rely on the initial orientation, or, more strongly, vertices are uniquely partitioned into rings, giving Property D1.

**Theorem 1** *The density decomposition is unique.*

**Proof:** The maximum indegree of two egalitarian orientations for a given graph is the same [9, 3, 35]. Suppose, for a contradiction, that there are two egalitarian orientations (red and blue) for $G$, resulting in density decompositions $R_0, R_1, \ldots R_k$ and $B_0, B_1, \ldots B_k$, respectively. Let $i$ be the largest index such that $R_i \neq B_i$.

We compare the orientation of the edges with one endpoint in $S = R_i \setminus B_i$ between the two orientations (illustrated in Figure 2). Since the orientations are egalitarian:

1. All the edges between $B_i$ and $S$ are directed into $S$ in the blue orientation.
2. All the edges between $S$ and $\{\cup_{j=i+1}^{k} R_j\} \setminus S$ are directed into $S$ with respect to both red and blue orientations.
3. All edges between $S$ and $\{\cup_{j=0}^{i-1} R_j\} \setminus S$ are directed out of $S$ with respect to the red orientation.

Based on these orientations, we have:

**Observation 1** *The number of edges directed into $S$ in the blue orientation is at least the number of edges directed into $S$ in the red orientation.*

We will show that $R_i \subseteq B_i$; symmetrically $B_i \subseteq R_i$, completing the theorem.

With respect to the blue orientation, all vertices in $S$ have indegree strictly less than $i$. Further, by the observation, the total indegree shared amongst the vertices in $S$ with respect to the red orientation is at most that of the blue orientation. Since all vertices in $S$ have indegree $i$ or $i-1$ with respect to the red orientation, and, by the observation, the total indegree shared amongst the vertices in $S$ with respect to the red orientation is at most that of the blue orientation, all vertices in $S$ have indegree $i-1$ with respect to the red orientation.

In order for every vertex in $S$ to have indegree $i-1$ in the red orientation, all vertices that are directed into $S$ in the blue orientation, must also be directed into $S$ in the red orientation; in particular this is true about the edges between $S$ and $R_i \setminus S$. Therefore, none of the vertices in $S$ (which have indegree $i-1$) reaches a vertex of $R_i \setminus S$ of indegree $i$ with respect to the red orientation. This contradicts the definition of $R_i$; therefore $S$ must be empty. □

The subgraph of a graph $G$ *induced* by a subset $S$ of the vertices of $G$ is defined as the set of vertices $S$ and the subset of edges of $G$ whose endpoints are both in $S$; we denote this by $G[S]$. We will show that both the densest subgraph and the subgraph induced by the vertices of highest rank have density between $k-1$ and $k$. Recall that $k$ is the maximum indegree of a vertex in an egalitarian orientation of $G$ and that $R_i$ is the set of vertices in the $i^{th}$ ring of the density decomposition. We will refer to $R_k$ as the densest ring.

Note that Property D2 has been proven before in another context. It follows from a theorem of Frank and Gyárfás [12] that if $\ell$ is the maximum outdegree in an orientation that minimizes the maximum outdegree then the density of the graph, $d$, is such that $\lceil d \rceil \leq \ell$. We use the following two lemmas to prove Property D2. Recall that $k$ is the maximum index of a non-empty ring.

**Lemma 1** *The density of the subgraph induced by the vertices in $R_k$ is in the range $(k-1, k]$.*

**Proof:** All vertices in $R_k$ have indegree $k$ or $k-1$ in $G$. Since any edge incident to a vertex in $R_k$ but not in $G[R_k]$ is directed out of $R_k$ in $G$, the indegree of every vertex in $G[R_k]$ is $k$ or $k-1$. Let $n_k$ be the number of vertices of indegree $k$ in $G[R_k]$ and $n_{k-1}$ be the number of vertices of degree $k-1$ in $G[R_{k-1}]$. Therefore, the number of edges in $G[R_k]$ is $kn_k + (k-1)n_{k-1}$ and:

$$\text{density}(G[R_k]) = \frac{kn_k + (k-1)n_{k-1}}{n_k + n_{k-1}} \leq k$$

Since there is at least one vertex of indegree $k$ in $G[R_k]$, $n_k > 0$. Therefore:

$$\frac{kn_k + (k-1)n_{k-1}}{n_k + n_{k-1}} = \frac{(k-1)(n_k + n_{k-1}) + n_k}{n_k + n_{k-1}} > k-1$$

□

**Lemma 2** *The density of the densest subgraph is in the range $(k-1, k]$, where $k$ the maximum index of a non-empty ring.*

**Proof:** Let $H$ be the densest subgraph and let each edge of $H$ inherit the orientation of the same edge in an egalitarian orientation of $G$. Every vertex of $H$ has indegree at most $k$ (when restricted to $H$). Therefore

$$\text{density}(H) \leq \frac{n_H k}{n_h} \leq k$$

where $n_H$ is the number of vertices in $H$. Furthermore, by Lemma 1, the density of $G[R_k]$ is greater than $k-1$ and so the densest subgraph must be at least this dense. □

The upper bound given in Lemma 2 proves Property D2 of the density decomposition.

**Corollary 2** *The subgraph induced by the vertices of $R_k$ is at least as dense as the density of the density subgraph less one.*

The following theorem relies on the fact that the density decomposition is unique and proves Property D3.

**Theorem 3** *The densest subgraph of a graph $G$ is induced by a subset of the vertices in the densest ring of $G$.*

**Proof:** First note that the densest subgraph is an induced subgraph, for otherwise, the subgraph would be avoiding including edges that would strictly increase the density. Let $S$ be a set of vertices that induces a densest subgraph of $G$. Consider a density decomposition of $G$ and let $k$ be the maximum rank of a vertex in $G$. Let $S_k = S \cap R_k$ and let $\bar{S}_k = S \setminus S_k$.

Let $A$ be the set of edges in $G[S_k]$, let $C$ be the set of edges in $G[\bar{S}_k]$, and let $B$ be the edges of $G[S]$ that are neither in $G[S_k]$ or $G[\bar{S}_k]$. We get

$$|B| + |C| \leq (k-1)|\bar{S}_k| \tag{1}$$

because all the edges in $B$ and $C$ have endpoints in $\bar{S}_k$ and all the vertices in $\bar{S}_k$ have indegree at most $k-1$ in the egalitarian orientation of $G$.

$$\text{density}(G[S]) = \frac{|A| + |B| + |C|}{|S_k| + |\bar{S}_k|} \tag{2}$$

$$\begin{aligned}
\text{density}(G[S_k]) &= \frac{|A|}{|S_k|} \\
&= \frac{\text{density}(G[S])(|S_k| + |\bar{S}_k|) - (|B| + |C|)}{|S_k|} \\
&\qquad \text{using Equation (2) to replace the numerator} \\
&= \text{density}(G[S]) + \frac{\text{density}(G[S])|\bar{S}_k| - (|B| + |C|)}{|S_k|} \\
&\geq \text{density}(G[S]) + \frac{\text{density}(G[S])|\bar{S}_k| - (k-1)|\bar{S}_k|}{|S_k|} \\
&\qquad \text{by Inequality (1)} \\
&> \text{density}(G[S]) + \frac{(k-1)|\bar{S}_k| - (k-1)|\bar{S}_k|}{|S_k|} \\
&\qquad \text{by Lemma 2} \\
&= \text{density}(G[S])
\end{aligned}$$

Therefore, removing the vertices of $G[S]$ that are not in $R_k$ produces a graph of strictly greater density. □

Note that there are indeed cases where the densest subgraph is induced by a strict subset of vertices in the top ring. See Figure 3 for an example.

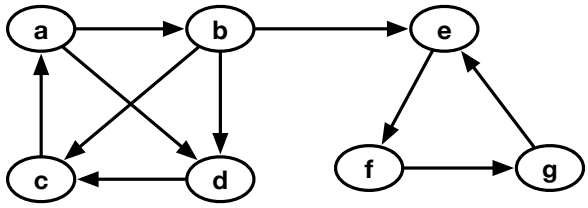


Figure 3: The orientation shown is an egalitarian orientation. In this graph all vertices are in the top ring. However, only the vertices in $K_4$, $a, b, c$ and $d$ are in the densest subgraph.

## 2.2 Interpretation of density rank

We can interpret orientations as assigning responsibility: if an edge is oriented from vertex $a$ to vertex $b$, we can view vertex $b$ as being *responsible* for that connection. Indeed several allocation problems are modelled this way [9, 2, 35, 3, 16]. Put another way, we can view a vertex as wishing to shirk as many of its duties (modelled by incident edges) by assigning these duties to its neighbors (by orienting the linking edge away from itself). Of course, every vertex wishes to shirk as many of its duties as possible. However, the topology of the graph may

prevent a vertex from shirking too many of its duties. In fact, the egalitarian orientation is the assignment in which every vertex is allowed to simultaneously shirk as many duties as allowed by the topology of the graph. An example is given in Figure 4; although vertices $a$ and $b$ both have degree 7, in the star graph (left) $a$ can shirk all of its duties, but in the clique graph (right) $b$ can only shirk half of its duties. There is a clear difference between these two cases that is captured by the density rank of $a$ and $b$ that is not captured by the degree of $a$ and $b$. For example, if these were co-authorship graphs, the star graph may represent a graph in which author $a$ only co-authors papers with authors who never work with anyone else whereas the clique graph shows that author $b$ co-authors with authors who also collaborate with others. One may surmise that the work of author $a$ is more reliable or respected than the work of author $b$.

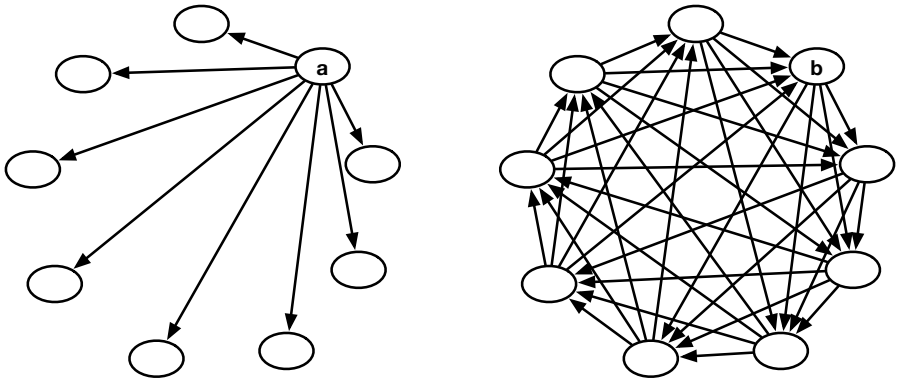


Figure 4: Two egalitarian orientations for graphs with 9 vertices. This example generalizes to any number of vertices (Theorem 4).

**Theorem 4** *For a clique on $n$ vertices, there is an orientation where each vertex has indegree either $\lfloor n/2 \rfloor$ or $\lfloor n/2 \rfloor - 1$.*

**Proof:** Give the vertices of the clique an ordering, $v_1, v_2, \ldots v_n$. Orient the edges between $v_1$ and $v_2, \ldots, v_{\lfloor n/2 \rfloor + 1}$ toward $v_1$ and edges between $v_1$ and $v_{\lfloor n/2 \rfloor + 2}, \ldots v_n$ toward $v_{\lfloor n/2 \rfloor + 2}, \ldots v_n$. Clearly $v_1$ has indegree $\lfloor n/2 \rfloor$. Similarly, for $v_2$: Orient the edges between $v_2$ and $v_3, \ldots, v_{\lfloor n/2 \rfloor + 2}$ toward $v_2$ and edges between $v_2$ and $v_{\lfloor n/2 \rfloor + 3}, \ldots v_n$ toward $v_{\lfloor n/2 \rfloor + 3}, \ldots v_n$. Clearly $v_2$ has indegree $\lfloor n/2 \rfloor$. Continue in this fashion until $v_n$. It is immediate that $v_1, v_2, \ldots v_{\lfloor n/2 \rfloor}$ have indegree $\lfloor n/2 \rfloor$. Now for the remaining vertices: Consider $v_i$, $\lfloor n/2 \rfloor < i \leq n$. $v_i$ has $n - i$ incoming edges from vertices $v_{i+1}, \ldots v_n$ and also $i - \lfloor n/2 \rfloor - 1$ incoming edges from $v_1, \ldots, v_{i - \lfloor n/2 \rfloor - 1}$. Therefore $v_i$ has indegree $\lfloor n/2 \rfloor - 1$. Therefore all vertices in the clique have indegree $\lfloor n/2 \rfloor$ or $\lfloor n/2 \rfloor - 1$. Clearly such an orientation is egalitarian. □

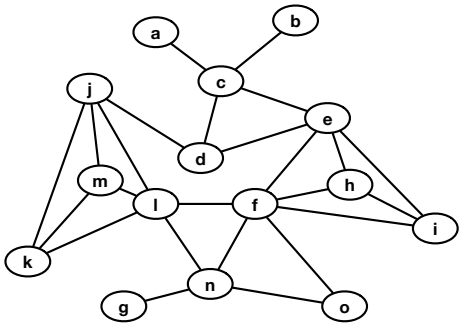


Figure 5: The top ring contains vertices $c, d, e, f, h, i, k, l, m, n$ and $o$ but $c, d, n$, and $o$ are not in the top core. The density of the subgraph induced by the top ring is $21/12 = 1.75$ while the density of the subgraph induced by the top core is $13/8 = 1.625$.

## 2.3 Relationship to $k$-cores

A $k$-core of a graph is the maximal subgraph whose vertices all have degree at least $k$ [33]. A $k$-core is found by repeatedly deleting vertices of degree less than $k$ while possible. For increasing values of $k$, the $k$-cores form a nesting hierarchy (akin to our density decomposition) of subgraphs $H_0, H_1, \ldots, H_p$ where $H_i$ is an $i$-core and $p$ is the smallest integer such that $G$ has an empty $(p+1)$-core. For graphs generated by the $G_{n,p}$ model, most vertices are in the $p$-core [22, 30] For the preferential attachment model, all vertices except the initial vertices belong to the $c$-core, where $c$ is the number of edges connecting to each new vertex [1].

These observations are similar to those we find for the density distribution (Section 3) and many of the observations we make regarding the similarity of the degree and density distributions of real-world graphs also hold for $k$-core decompositions [27]. However $k$-cores are defined by minimum induced degrees and so are only indirectly related to density. We make formal this much looser connection to density than the density decomposition in Lemma 3. See Figure 5 for an example of where the density of the top core is less then the density of the top ring. Further, while the core decomposition of a graph can be found in time linear in the number of edges [23, 6, 10] as opposed to the quadratic time required for the density decomposition [9], core decompositions do not lend themselves to a framework for building synthetic graphs, since it is not clear how to generate a $p$-core at random, whereas density decompositions do (Section 4).

Recall that identifying the vertices in $\cup_{j>i} R_j$ and deleting the vertices in $\cup_{j<i} R_j$ leaves a graph $G$ whose density is in the range $(i-1, i]$ (for $|R_i|$ sufficiently large). We find that the bound on density for the corresponding cores is much looser.

**Lemma 3** *Given a core decomposition $H_0, H_1, \ldots, H_k$ of a graph, the subgraph formed by identifying the vertices in $\cup_{j>i} H_j$ and deleting the vertices in $\cup_{j<i} H_j$*

*has density in the range* $[\frac{i}{2}, i)$ *for* $|H_i|$ *sufficiently large.*

**Proof:** Let $n$ be the number of vertices in the described subgraph: $n = |H_i|+1$. Let $d$ be the degree of the vertex resulting from the identification of $\cup_{j>i} H_j$. Since every vertex in $H_i$ has degree at least $i$ in the subgraph, the density of the subgraph is at most $\frac{\frac{1}{2}(i \cdot n + d)}{n}$, from which the lower bound of the lemma follows since $d > 0$. This lower bound is also tight when $H_i$ induces an $i$-regular graph.

Further, the $i$-core is witnessed by iteratively deleting vertices of degree at most $i$ while such vertices exist. The subgraph will have the greatest density (the most edges) if each deletion removes a vertex of degree exactly $i$. Then the subgraph has density at most $\frac{i \cdot (n-1)}{n}$. □

# 3 The similarity of degree and density distributions

In this section, we consider ten varied networks (see Table 1). We compare the normalized density and normalized degree distributions of these networks. We find our results to be consistent across biological, technical, and social networks.

The normalized density $\boldsymbol{\rho}$ and degree $\boldsymbol{\delta}$ distributions for three networks (AS 2013, PHYS 2005, and DBLP) are given in Figure 6, illustrating the similarity of the distributions. We quantify the similarity between the density and degree distributions of these networks using the Bhattacharyya coefficient, $\beta$ [7] (note that we could use any similar distance metric for two probability distributions). For two normalized $\mathbf{p}$ and $\mathbf{q}$, the Bhattacharyya coefficient is:

$$\beta(\mathbf{p}, \mathbf{q}) = \sum_i \sqrt{p_i \cdot q_i}.$$

$\beta(\mathbf{p}, \mathbf{q}) \in [0, 1]$ for normalized, positive distributions; $\beta(\mathbf{p}, \mathbf{q}) = 0$ if and only if $\mathbf{p}$ and $\mathbf{q}$ are disjoint; $\beta(\mathbf{p}, \mathbf{q}) = 1$ if and only if $\mathbf{p} = \mathbf{q}$. We denote the Bhattacharyya coefficient comparing the normalized density $\boldsymbol{\rho}$ and degree $\boldsymbol{\delta}$ distributions, $\beta(\boldsymbol{\rho}, \boldsymbol{\delta})$ for a graph $G$ by $\beta_{\rho\delta}(G)$. Specifically,

$$\beta_{\rho\delta}(G) = \beta(\boldsymbol{\rho}, \boldsymbol{\delta}) = \sum_i \sqrt{\rho_i \cdot \delta_i}$$

where $\rho_i$ is the fraction of vertices in the $i^{th}$ ring of the density decomposition of $G$ and $\delta_i$ is the fraction of vertices of *total* degree $i$ in $G$; we take $\rho_i = 0$ for $i > k$ where $k$ is the maximum ring index. Refer to Figure 7. For all the networks in our data set, $\beta_{\rho\delta} > 0.78$. Note that if we exclude the Gnutella and Amazon networks, $\beta_{\rho\delta} > 0.9$. We point out that the other networks are self-determining in that each relationship is determined by at least one of the parties involved. On the other hand, the Gnutella network is highly structured and designed and the Amazon network is a is a one-mode projection of the buyer-product network (which is in turn self-determining).

Self-determining networks

| Name | Vertices | # Vertices | Edges | # Edges | Source |
|---|---|---|---|---|---|
| AS | autonomous systems | 44,729 | routing agreements | 170,735 | [42] |
| DBLP | computer scientists | 317,080 | at least one co-authored paper | 1,049,866 | [41] |
| Enron | email addresses | 36,692 | at least one email exchanged | 183,831 | [18] |
| Epinions | `epinions.com` members | 75,879 | self-indicated trust | 405,740 | [34] |
| Facebook | Facebook user | 4,039 | Facebook friends | 88234 | [24] |
| PHYS | condensed matter physicists | 40,421 | at least one co-authored paper | 175,692 | [28] |
| Slashdot | `slashdot.org` members | 82,168 | indication of friend or foe | 504,230 | [21] |
| Wikivote | `wikipedia.org` users | 7,115 | votes for administrator role | 103,689 | [20] |

Non-self-determining networks

| Name | Vertices | # Vertices | Edges | # Edges | Source |
|---|---|---|---|---|---|
| Amazon | products | 334,863 | pairs of frequently co-purchased items | 925,872 | [41] |
| Gnutella | network hosts | 22,687 | connections for file sharing | 54,705 | [31] |

Table 1: Network data sets. For naturally directed networks (Enron, Epinions and Wikivote), we ignore the directions and study the underlying undirected network. We likewise ignore edge annotations (e.g. friend or foe in the Slashdot network). We use three snapshots of the AS network (from 1999, 2005 and 2011) and three snapshots of the PHYS network (for papers posted to `arxiv.org` prior to 1999, 2003 and 2005). Note that the structure of the Gnutella network is given by external system design specifications.

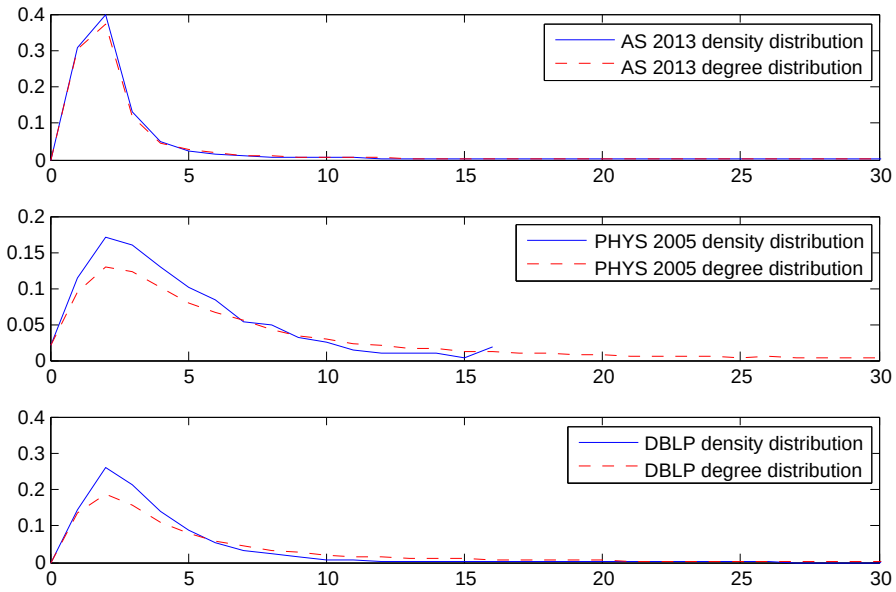


Figure 6: In the AS network vertices represent autonomous systems and two autonomous systems are connected if there is a routing agreement between them [42]. In the PHYS network vertices represent condensed matter physicists and two physicists are connected if they have at least one co-authored paper [28]. In the DBLP network, vertices represent computer scientists and two computer scientists are connected if they have at least one co-authored paper [41]. The (truncated) normalized density and degree distributions are displayed. The degree distributions have long diminishing tails. AS 2013 has 67 non-empty rings, but rings 31 through 66 contain less than 1.5% of the vertices; ring 67 contains 0.75% of the vertices. DBLP has 4 non-empty rings denser than ring 30 that are disconnected; rings 32, 40, 52 and 58 contain 0.02%, 0.01%, 0.03% and 0.04% of the vertices, respectively.

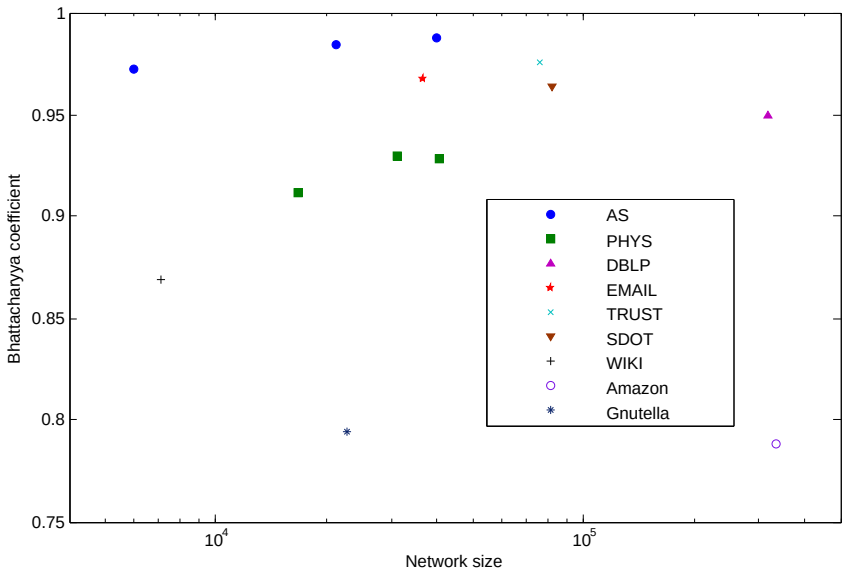


Figure 7: Similarity ($\beta_{\rho\delta}$) of density and degree distributions for 9 diverse networks. We introduced AS, PHYS, and DBLP in Figure 6. In the EMAIL network vertices represent Enron email addresses and two addresses are connected if there has been at least one email exchanged between them [18]. In the TRUST network vertices represent `epinions.com` members and two members are connected if one trusts the other [34]. In the SDOT network vertices represent `slashdot.org` members and two members are connected if they are friends or foes [21]. In the WIKI network vertices represent `wikipedia.org` users and two users are connected if one has voted for the other to be in an administrative role [20]. In the Amazon network vertices represent products and two products are connected if they are frequently purchased together [41]. In the Gnutella network vertices represent network hosts and two hosts are connected if they share files [31]. EMAIL , TRUST, and WIKI are naturally directed networks. For these networks, we ignore direction and study the underlying undirected networks. Notice that both the Amazon and Gnutella networks are highly structured. It is not surprising that these networks would have a weaker connection between the density and degree distributions.

Perhaps this is not surprising, given the close relationship between density and degree; one may posit that the density distribution $\boldsymbol{\rho}$ simply bins the degree distribution $\boldsymbol{\delta}$. However, note that a vertex's degree is its *total* degree in the undirected graph, whereas a vertex's rank is within one of its *indegree* in an egalitarian orientation. Since the total indegree to be shared amongst all the vertices is half the total degree of the graph, we might assume that, if the density distribution is a binning of the degree distribution, the density rank of a vertex of degree $d$ would be roughly $d/2$. That is, we may expect that the density distribution is halved in range and doubled in magnitude ($\rho_i \approx 2\delta_{2i}$). If this is the case, then

$$\beta(\boldsymbol{\rho}, \boldsymbol{\delta}) \approx \sum_d \sqrt{\rho_i \delta_i} \approx \sum_d \sqrt{2\delta_d \delta_{2d}}.$$

If we additionally assume that our graph has a power-law degree distribution such as $\delta_x \propto 1/x^3$,

$$\beta(\boldsymbol{\rho}, \boldsymbol{\delta}) \approx \int_1^\infty \sqrt{\frac{2}{x^3}\left(2\frac{2}{(2x)^3}\right)}\mathrm{d}x = 0.5$$

(after normalizing the distributions and using a continuous approximation of $\beta$). Even with these idealized assumptions, this does not come close to explaining $\beta_{\rho\delta}$ being in excess of 0.78 for the networks in our data set. Further to that, for many synthetic networks $\beta_{\rho\delta}$ is close to 0, as we discuss in the next section. We note that this separation between similarities of density and degree distributions for the empirical networks and synthetic networks can be illustrated with almost any divergence or similarity measure for a pair of distributions.

## 3.1 The dissimilarity of degree and density distributions of random networks

In contrast to the measurably similar degree and density distributions of real-world networks, the degree and density distributions are measurably *dis*similar for networks produced by many common random network models; including the preferential attachment (PA) model of Barabasi and Albert [5] and the small world (SW) model of Watts and Strogatz [37]. We will discuss the degree-sequence model in Section 4.3. We use $\tilde{\beta}_{\rho\delta}(M)$ to denote the Bhattacharyya coefficient comparing the expected degree and density distributions of a network generated by a model $M$.

**Preferential attachment networks** In the PA model, a small number, $n_0$, of vertices seed the network and vertices are added iteratively, each attaching to a fixed number, $c$, of existing vertices. Consider the orientation where each added edge is directed toward the newly added vertex; in the resulting orientation, all but the $n_0$ seed vertices have indegree $c$ and the maximum indegree is $c$. At most $cn_0$ path reversals will make this orientation egalitarian, and, since $cn_0$ is typically very small compared to $n$ (the total number of vertices), most

of the vertices will remain in the densest ring $R_c$. Therefore PA networks have nearly-trivial density distributions: $\rho_c \approx 1$. On the other hand the expected fraction of degree $c$ vertices is $\delta_c \approx 2/c^3$. Therefore $\tilde{\beta}_{\rho\delta}(\text{PA}) \approx \sqrt{2/c^3}$, which quickly approaches 0 as $c$ grows, and is $\leq 0.25$ for $c \geq 2$.

**Small-world networks** A small-world network is one generated from a $d$-regular graph[2] by reconnecting (uniformly at random) at least one endpoint of every edge with some probability. For probabilities close to 0, a network generated in this way is close to $d$-regular; for probabilities close to 1, a network generated this way approaches one generated by the random-network model ($G_{n,p}$) of Erdös and Rényi [11]. In the first extreme, $\tilde{\beta}_{\rho\delta}(SW) = 0$ (Lemma 4 below) because all the vertices have the same degree and the same rank. As the reconnection probability increases, vertices are not very likely to change rank while the degree distribution spreads slightly. In the second extreme, the highest rank of a vertex is $\lfloor c/2 \rfloor + 1$ [40] and, using an observation of the expected size of the densest subnetwork [26], with high probability nearly all the vertices have this rank. It follows that

$$\tilde{\beta}_{\rho\delta}(G_{n,p}) \approx \sqrt{\frac{c^{c/2}}{e^{-c}(c/2)!}},$$

which approaches 0 very quickly as $c$ grows. We verified this experimentally finding that $\tilde{\beta}_{\rho\delta}(G_{n,p}) < 0.5$ for $c \geq 5$.

**Lemma 4** *For $d \geq 3$, $\beta_{\rho\delta}(G) = 0$ for any $d$-regular graph $G$ with $d \geq 3$.*

**Proof:** We argue that $\rho_d = 0$, proving the lemma since $\delta_d = 1$ for a $d$-regular graph. For a contradiction, suppose $\rho_d > 0$. Then $|R_d| = x$ for some $x > 0$, where $R_d$ is the set of vertices of $G$ in the $d^{th}$ ring of $G$'s density decomposition. Note that the highest rank vertex in $G$ has rank at most $d$, since there are no vertices with degree $> d$. Let $H$ be the subgraph of $H$ containing all the vertices of $R_d$ and all the edges of $G$ both of whose endpoints are in $R_d$. $H$ has at least one vertex of indegree $d$ and all other vertices have indegree at least $d-1$; therefore $H$ must have at least $d + (x-1)(d-1)$ edges. On the other hand, the total degree of every vertex in $H$ is at most $d$, so $H$ has at most $dx/2$ edges. We must have $d + (x-1)(d-1) \leq dx/2$, which is a contradiction for $d \geq 3$ and $x > 0$. □

Note that there are many other random networks models that we could have analyzed. Our results are not intended to be exhaustive, rather they are simply intended to be interesting observations.

[2] A graph in which every vertex has degree $d$.

# 4 Random networks with given density distributions

Motivated by our observations from Section 3.1, that many random network models produce networks with dissimilar degree and density distributions, we present a random network model based on the density distribution with an aim to achieve a realistic similarity between the degree and density distributions as observed in Section 3.

Given a density distribution $\boldsymbol{\rho}$, we can generate a network with $n$ vertices having this density distribution using the following *abstract model*:

**Input:** density distribution $\boldsymbol{\rho}$ and target size $n$
**Output:** an network $G$ with $n$ vertices and density distribution $\boldsymbol{\rho}$
1: Initialize $G$ to be a network with empty vertex set $V$
2: **for** $i = |\boldsymbol{\rho}|, \ldots, 0$ **do**
3: $R_i \leftarrow$ set of $\lfloor \rho_i n \rfloor$ vertices
4: add $R_i$ to $V$
5: **for** each vertex $v \in R_i$ **do**
6: connect $i$ vertices of $V$ to $v$

Using this generic model, we propose two specific models, the *random density distribution model* (RDD - Section 4.1) and the *hierarchical small worlds model* (HSW - Section 4.2), by specifying how the neighbors are selected in Step 6. First we show that this abstract model does indeed generate a network with the given density distribution:

**Lemma 5** *The network resulting from the abstract model has density distribution* $\boldsymbol{\rho}$.

**Proof:** We argue that the orientation given by, in Step 6, directing the added edges into $v$ is egalitarian. For a contradiction, suppose there is a reversible path. There must be an edge on this path from a vertex $x$ to a vertex $y$ such that the in-degree of $y$ is strictly greater than the degree of $x$. By construction, then, $x$ was added after $y$ and so an edge between $x$ and $y$ must oriented into $x$, contradicting the direction required by the reversible path.

Finally, since the vertices in set $R_i$ have indegree $i$ according to this orientation, the orientation is a witness to a density decomposition of the given distribution. □

Notice that in this construction, vertices in $R_i$ will have indegree $i$ while a network with the same density decomposition may have vertices in $R_i$ with indegree $i-1$. We could additionally specify the number of vertices in $R_i$ that have indegree $i$ and indegree $i-1$; this would additionally require ensuring that there is an egalitarian orientation in which all the vertices destined to have indegree $i-1$ in $R_i$ *reach* vertices of indegree $i$ in $R_i$. We believe this is needlessly over-complicated and, indeed, over-specification that will have little affect on the generation *large* realistic networks.

Further notice that this abstract model may generate a network that is not simple. Without further constraint, in Step 6, $v$ may connect to itself (introducing a self-loop) or to a vertex that $v$ is already connected to (introducing parallel edges). We adopt a simple technique used for generating $d$-regular graphs [25]: we constrain the choice in Step 6 to vertices of $V$ that are not $v$ itself nor neighbors of $v$. McKay and Wormald prove this constraint still allows for uniformity of sampling of $d$-regular graphs when $d$ is sufficiently small ($d = O(n^{1/3})$) [17]; likewise, since $i$ is small compared to $|R_i|$ for large networks, adopting this technique should not affect our sampling. In our two specific models, described below, we ensure the final network will be simple using this technique.

## 4.1 Random density distribution model

For the RDD model, we choose $i$ vertices from $V$ uniformly at random in Step 6. We use this to model four networks in our data set (AS, DBLP, EMAIL, and TRUST). For each given network, we generate another random network having the given network's number of vertices and density distribution. Although we are only specifying the distribution of the vertices over a density decomposition, the resulting degree distributions of the RDD networks are very similar to the original networks they are modeling. We use the Bhattacharyya coefficient to quantify the similarity between the normalized degree distribution of an RDD network and the normalized degree distribution of the original network; we denote this by $\beta_{\delta\delta}$ (to distinguish from our use of the Bhattacharyya coefficient to compare degree distributions to density distributions. For all four models, $\beta_{\delta\delta} > 0.93$ (Figure 8). Further, the average path lengths of the RDD networks are realistic, within 2 of the average path lengths of the original networks (Figure 9).

However, the clustering coefficients of the RDD networks are unrealistically low (Figures 8 and 9). Upon further inspection, we find that, for example, the PHYS networks have many more edges between vertices of a common ring of its density decomposition than between rings as compared to the corresponding RDD model. For the RDD model, we can compute the expected fraction of edges that will have one endpoint in $R_i$ and one endpoint in $R_j$. Since there are $|R_j||R_i|$ such edges to choose from (for $j > i$) and at most $|R_i|(\frac{1}{2}(|R_i| - 1) + \sum_{j>i} |R_j|)$ edges between $R_i$ and $R_j$ (for $j > i$), we would expect this fraction to be:

$$\frac{|R_j|}{\frac{1}{2}(|R_i| - 1) + \sum_{j>i} |R_j|} \text{ for } j > i \text{ and } \frac{\frac{1}{2}(|R_i| - 1)}{\frac{1}{2}(|R_i| - 1) + \sum_{j>i} |R_j|} \text{ for } i = j \quad (3)$$

In Figure 10 we plot the difference between the actual fraction of edges connecting $R_i$ to $R_j$ in the PHYS networks with this expected fraction for all values of $j - i$. We see that when $j - i = 0$, or for edges with both endpoints in the same ring, there is a substantially larger number of edges in the original networks than is being captured by our model. This provides one explanation for the low clustering coefficients produced by the RDD model.

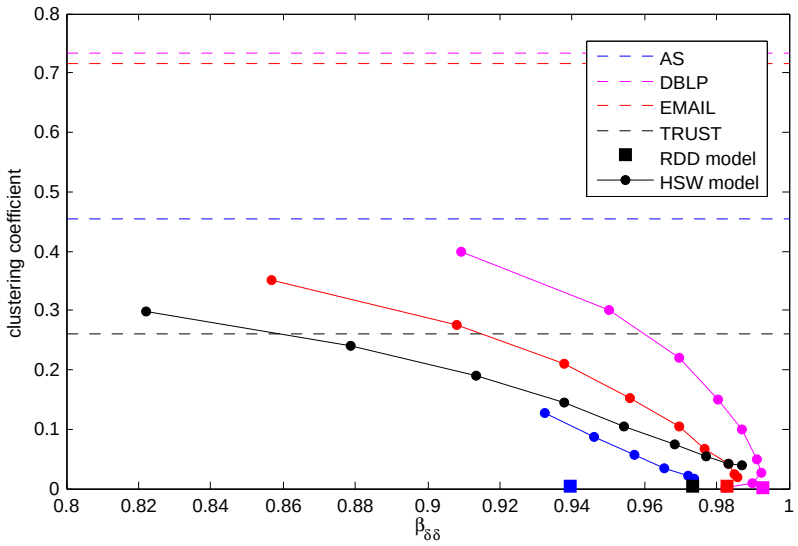


Figure 8: Clustering coefficient versus similarity of degree distribution (between models and original networks, $\beta_{\delta\delta}$) for RDD and HSW models. Measurements for the SW model networks are not shown as $\beta_{\delta\delta} < 0.4$ for all networks generated. Dotted lines represent the clustering coefficients for the original networks.

## 4.2 Hierarchical small worlds model

We provide a more sophisticated model which addresses the unrealistically low clustering coefficients of the RDD model by generating a small world (SW) network among the vertices of each ring of the density decomposition. Recall that a SW network on $n$ vertices, average degree $d$ and randomization $p$ network is created as follows: order the vertices cyclically and connect each vertex to the $d$ vertices prior to it; with probability $p$ reconnect one endpoint of each edge to another vertex chosen uniformly at random. The SW model provides a trade-off between clustering coefficient and average path length: as $p$ increases, the clustering coefficient and the average path length decreases [37].

In the hierarchical small worlds (HSW) model, for vertices in $R_i$, we create a SW network on $|R_i|$ vertices and average degree $i$ in the same way, except for how we reconnect each edge with probability $p$. For an edge $uv$ where $u$ is a vertex within $d$ vertices prior to $v$ in the cyclic order, we select a vertex $x$ uniformly at random from $\cup_{j>i} R_j$ and replace $uv$ with $xv$. For the densest ring, we select a vertex uniformly at random from the densest ring.

This process is exactly equivalent to the following: order $R_i$ cyclically; for each $v \in R_i$, with probability $p$, connect each of the $i$ vertices before $v$ in this order to $v$; if $c \leq i$ neighbors for $v$ are selected in this way, select $i - c$ vertices

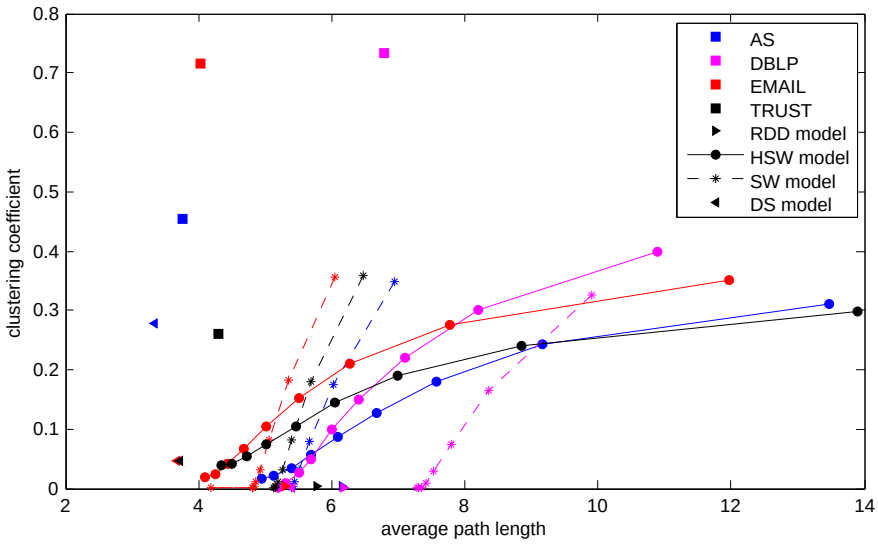


Figure 9: Clustering coefficient versus average path length for RDD, HSW, SW and DS models. Colors indicate the network being modelled. Squares denote the data for the original networks.

uniformly at random from $\cup_{j>i} R_j$ (or $R_i$ if this is the densest ring) and connect these to $v$. Clearly, this is a specification of neighbor selection for Step 6 of the abstract model.

For the AS, DBLP, EMAIL, and TRUST networks in our data set, we generate a random network according to the HSW model that is of the same size and density distribution of the original network. We do so for $p = 0.1, 0.2, \ldots, 0.9$. As with the SW model, the HSW model provides a similar trade-off between clustering coefficient and average path length (Figure 9), although the relationship is less strong. In addition, we observe a similar trade-off between $p$ and degree distribution: as $p$ increases, the degree distribution approaches that of the original network (Figure 8). This is in sharp contrast to the SW model which have degree distributions far from the original (normal vs. close to power law).

## 4.3 Comparing to the degree sequence model

We also compare our models (RDD and HSW) to a *degree sequence* (DS) model. For a given degree distribution or sequence (assignment of degree to each vertex), a DS model will generate a graph, randomly, having that degree sequence. We use the model of Viger and Latapy which generates a connected, simple graph

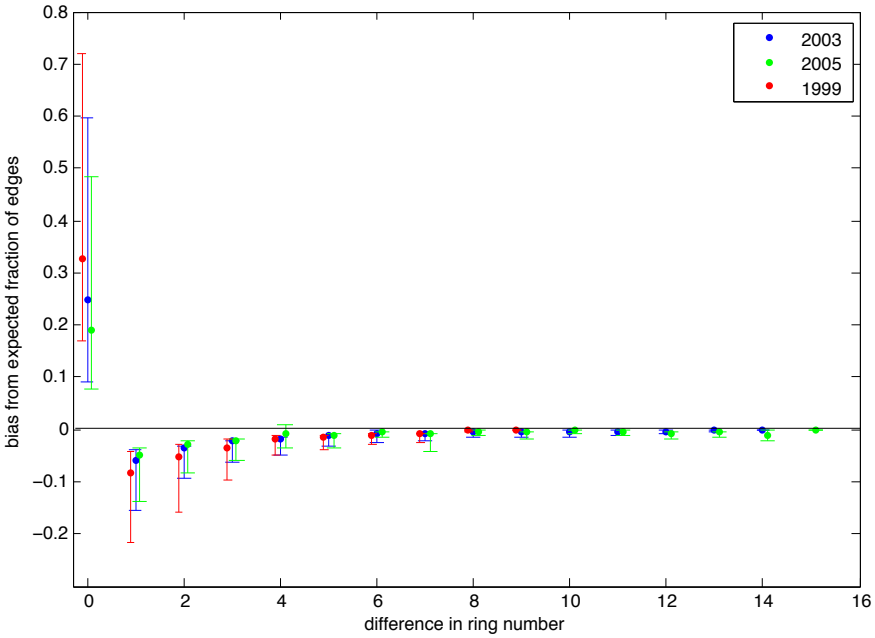


Figure 10: Range of difference between the actual fraction of edges connecting $R_i$ to $R_j$ and expected fraction (Equation (3)) over all $j \geq i$ as a function of $j - i$ for three PHYS networks. Error bars show max and min values of these differences and dots indicate the average.

by iteratively selecting neighbors for vertices (from highest remaining degree to be satisfied to lowest) and randomly shuffling to prevent the process from getting stuck (if no new neighbor exists that has not yet fulfilled its prescribed degree) [36]. As with RDD and HSW we generate a network using this DS model corresponding to the degree sequence of the AS, EMAIL, and TRUST networks. The clustering coefficients of the resulting networks are much lower than in the real-world networks (Figure 9); in the case of the AS network, this mismatch is less extreme, most likely because this network has an extremely long tail with a vertex with degree 4,171; many vertices would connect to these high degree vertices, providing an opportunity for clustering. The average path lengths are close to the original networks. Notably, the density distributions of the networks generated by the DS model are very similar to their degree distribution, all having $\beta_{\rho\delta} > 0.9$.

These observations for the DS model add evidence to our proposal that in order to generate realistic networks, one must distinguish between types of vertices; doing so results in networks that resemble real-world networks. However, we must note that the DS model suffers from two drawbacks. First, the al-

gorithms for generating such networks are much less efficient than our models (RDD and HSW, which run in linear time); in order to guarantee simplicity and connectivity, the reshuffling required incurs a large computational overhead, particularly when the degree sequence includes very high degree vertices (such as in the AS network). Second, the amount of information required to specify network generation via the DS model is an order of magnitude greater than our abstract model. In the former, the degree of every vertex must be specified, or at least the number of vertices having each degree. For example, the SLASH network has 457 unique degrees (and a maximum degree of 2553) while only having 61 non-empty rings in the density decomposition.

### 4.4 Conclusion

We close by pointing out that the abstract model as presented at the start of Section 4 is very flexible. One may specify any number of ways to choose how neighbors are selected in Step 6. As an additional example, one may select neighbors with probability proportional to their current degree as in the preferential attachment model; this would likely result in lower average path lengths, but also unrealistically low clustering coefficients. Or, one could modify our HSW model by reconnecting to vertices in a preferential way; that is one could combine the SW and PA model within our abstract model. More than likely, different types of networks, such as autonomous system networks versus social networks, would be best modeled by different specifications of the abstract model. Needless to say, the most important quality that we believe our model provides is a realistic partitioning of the vertices into classes.

## 5 Acknowledgements

We would like to express our sincerest thanks to our reviewer and editor for helping us to make this paper strong and clear.

## References

[1] J. I. Alvarez-Hamelin. K-core decomposition of internet graphs: hierarchies, self-similarity and measurement biases, 2008. doi:10.3934/nhm.2008.3.371.

[2] Y. Asahiro, J. Jansson, E. Miyano, and H. Ono. Upper and lower degree bounded graph orientation with minimum penalty. In *Proceedings in Computing: The Australasian Theory Symposium*, pages 139–146, 2012.

[3] Y. Asahiro, E. Miyano, H. Ono, and K. Zenmyo. Graph orientation algorithms to minimize the maximum outdegree. *International Journal of Foundations of Computer Science*, 18(02):197–215, 2007. doi:10.1142/S0129054107004644.

[4] B. Bahmani, R. Kumar, and S. Vassilvitskii. Densest subgraph in streaming and mapreduce. *Proc. VLDB Endow.*, 5(5):454–465, Jan. 2012. doi:10.14778/2140436.2140442.

[5] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. doi:10.1126/science.286.5439.509.

[6] V. Batagelj and M. Zaversnik. An $O(m)$ algorithm for cores decomposition of networks. *Advances of Data Analysis and Classification*, 5:129–145, 2011.

[7] A. K. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.

[8] T. Biedl, T. Chan, Y. Ganjali, M. T. Hajiaghayi, and D. R. Wood. Balanced vertex-orderings of graphs. *Discrete Appl. Math.*, 148:27–48, April 2005. doi:10.1016/j.dam.2004.12.001.

[9] G. Borradaile, J. Iglesias, T. Migler, A. Ochoa, G. Wilfong, and L. Zhang. Egalitarian graph orientations. In *Journal of Graph Algorithms and Applications*, volume 21, pages 687–708, 2017. doi:10.7155/jgaa.00435.

[10] M. Charikar. Greedy approximation algorithms for finding dense components in a graph. In *Proceedings of the Third International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 84–95, London, UK, 2000. Springer-Verlag. URL: http://dl.acm.org/citation.cfm?id=646688.702972.

[11] P. Erdös and A. Rényi. On random graphs, I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959. URL: http://www.renyi.hu/~p_erdos/Erdos.html#1959-11.

[12] A. Frank and A. Gyárfás. How to orient the edges of a graph? *Colloquia Mathematica Societatis János Bolyai*, 1:353–364, 1976.

[13] H. d. Fraysseix and P. O. de Mendez. Regular orientations, arboricity, and augmentation. In *Proceedings of the DIMACS International Workshop on Graph Drawing*, pages 111–118. Springer-Verlag, 1995. doi:10.1007/3-540-58950-3_362.

[14] D. Gibson, R. Kumar, and A. Tomkins. Discovering large dense subgraphs in massive graphs. In *Proceedings of the 31st international conference on very large data bases*, pages 721–732. VLDB Endowment, 2005. URL: http://dl.acm.org/citation.cfm?id=1083592.1083676.

[15] A. Goldberg. Finding a maximum density subgraph. Technical report, University of California at Berkeley, Berkeley, CA, USA, 1984.

[16] N. J. A. Harvey, R. E. Ladner, L. Lovász, and T. Tamir. Semi-matchings for bipartite graphs and load balancing. *Journal of Algorithms*, 59:53–78, 2006. doi:10.1016/j.jalgor.2005.01.003.

[17] J. H. Kim and V. H. Vu. Generating random regular graphs. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, STOC '03, pages 213–222, New York, NY, USA, 2003. ACM. doi:10.1145/780542.780576.

[18] B. Klimt and Y. Yang. Introducing the Enron Corpus. In *First Conference on Email and Anti-Spam*, 2004.

[19] W. F. Klostermeyer. Pushing vertices and orienting edges. *Ars Combinatorial*, 51:65–75, 1999.

[20] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1361–1370, New York, NY, USA, 2010. ACM. http://snap.stanford.edu/data/. doi:10.1145/1753326.1753532.

[21] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *CoRR*, abs/0810.1355, 2008. http://snap.stanford.edu/data/. URL: http://arxiv.org/abs/0810.1355.

[22] T. Łuczak. Size and connectivity of the k-core of a random graph. *Discrete Mathematics*, 91(1):61–68, Aug. 1991. doi:10.1016/0012-365X(91)90162-U.

[23] D. W. Matula and L. L. Beck. Smallest-last ordering and clustering and graph coloring algorithms. *J. ACM*, 30(3):417–427, July 1983. doi:10.1145/2402.322385.

[24] J. Mcauley and J. Leskovec. Discovering social circles in ego networks. *ACM Transactions on Knowledge Discovery from Data*, 8(1):4:1–4:28, Feb. 2014. doi:10.1145/2556612.

[25] B. D. McKay and N. C. Wormald. Uniform generation of random regular graphs of moderate degree. *J. Algorithms*, 11(1):52–67, Feb. 1990. doi:10.1016/0196-6774(90)90029-E.

[26] A. Mehrabian. e-mail exchange between Glencora Borradaile and Abbas Mehrabian. e-mail, August 15, 2013.

[27] T. Migler. *The Density Signature*. PhD thesis, Oregon State University, 2014.

[28] M. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69:066133, Jun 2004. http://www-personal.umich.edu/~mejn/netdata/. doi:10.1103/PhysRevE.69.066133.

[29] M. Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010.

[30] B. Pittel, J. Spencer, and N. Wormald. Sudden emergence of a giant k-core in a random graph. *J. Comb. Theory Ser. B*, 67(1):111–151, May 1996. doi:10.1006/jctb.1996.0036.

[31] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal*, 6:2002, 2002. http://snap.stanford.edu/data/.

[32] B. Saha, A. Hoch, S. Khuller, L. Raschid, and X.-N. Zhang. Dense subgraphs with restrictions and applications to gene annotation graphs. In *Proceedings of the 14th Annual international conference on Research in Computational Molecular Biology*, RECOMB'10, pages 456–472, Berlin, Heidelberg, 2010. Springer-Verlag. doi:10.1007/978-3-642-12683-3_30.

[33] S. Seidman. Network structure and minimum degree. *Social Networks*, 5(3):269–287, 1983. doi:10.1016/0378-8733(83)90028-X.

[34] M. Tahajod, A. Iranmehr, and N. Khozooyi. Trust management for semantic web. In *Computer and Electrical Engineering, 2009. ICCEE '09. Second International Conference*, volume 2, pages 3–6, 2009. http://snap.stanford.edu/data/. doi:10.1109/ICCEE.2009.241.

[35] V. Venkateswaran. Minimizing maximum indegree. *Discrete Appl. Math.*, 143:374–378, September 2004. doi:10.1016/j.dam.2003.07.007.

[36] F. Viger and M. Latapy. Efficient and simple generation of random simple connected graphs with prescribed degree sequence. In *The Eleventh International Computing and Combinatorics Conference*, pages 440–449. Springer, 2005. doi:10.1007/11533719_45.

[37] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):409–10, 1998. doi:10.1038/30918.

[38] W. Wimmer. Ein Verfahren zur Verhinderung von Verklemmungen in Vermittlernetzen. http://www.worldcat.org/title/verfahren-zur-verhinderung-von-verklemmungen-in-vermittlernetzen/, October 1978.

[39] V. Wittorff. Implementation of constraints to ensure deadlock avoidance in networks, 2009. US Patent # 7,532,584 B2.

[40] P. G. Xavier Pérez-Giménez and C. Sato. Arboricity and spanning-tree packing in random graphs with an application to load balancing. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pages 317–326. SIAM, 2014.

[41] J. Yang and J. Leskovec. Defining and evaluating network communities based on ground-truth. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, MDS '12, pages 3:1–3:8, New York, NY, USA, 2012. ACM. http://snap.stanford.edu/data/. doi:10.1145/2350190.2350193.

[42] Y. Zhang. Internet AS-level Topology Archive. http://irl.cs.ucla.edu/topology/.