



HAL
open science

Small world networks and clustered small world networks with random connectivity

Faraz Zaidi

► **To cite this version:**

Faraz Zaidi. Small world networks and clustered small world networks with random connectivity. Social Network Analysis and Mining, 2012. hal-00679660

HAL Id: hal-00679660

<https://hal.science/hal-00679660>

Submitted on 16 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Small World Networks and Clustered Small World Networks with Random Connectivity

Faraz Zaidi

Received: date / Accepted: date

Abstract The discovery of small world properties in real world networks has revolutionized the way we analyze and study real world systems. Mathematicians and Physicists in particular have closely studied and developed several models to artificially generate networks with small world properties. The classical algorithms to produce these graphs artificially, make use of the fact that with the introduction of some randomness in ordered graphs, small world graphs can be produced. In this paper, we present a novel algorithm to generate graphs with small world properties based on the idea that with the introduction of some order in a random graph, small world graphs can be generated. Our model starts with a randomly generate graph. We then replace each node of the random graph with cliques of different sizes. This ensures that the connectivity between the cliques is random but the clustering coefficient increases to a desired level. We further extend this model to incorporate the property of community structures (clusters) found readily in real world networks such as social, biological and technological networks. These community structures are densely connected regions of nodes in a network that are loosely connected to each other. The model generates these clustered small world graphs by replacing nodes in the random graph with densely connected set of nodes. Experimentation shows that these two models generate small world and clustered small world graphs respectively as we were able to produce the desired properties of a small world network with high clustering coefficient and low average path lengths in both cases. Furthermore, we also calculated Relative Density and Modularity to show that the clustered networks indeed had community structures.

Keywords Small World Networks · Random Networks · Community Structures · Clustered Networks · Network Generation Models

F. Zaidi
Karachi Institute of Economics and Technology (KIET)
Karachi, Pakistan and
INRIA Bordeaux - Sud Ouest ,
Bordeaux, France
E-mail: faraz@pafkiet.edu.pk

1 Introduction

Small world networks can be readily found in many real world systems such as social networks [46,38], economic networks [17], transportation systems [13,35], epidemic spreading [31], metabolic networks [16,3], food web [22] and so on.

The discovery of small world properties has enabled researchers to understand, improve and manipulate these networks for useful applications. As an example, the spread of disease or epidemics has been well studied in social networks and small world properties have been used to devise strategies to isolate and quarantine infected communities [23]. Similarly Online social communities [34], Stock affiliation [41], air transportation [7], metabolic networks [43] are all examples where small world properties have been very useful to analyze these networks and enhance decision making processes.

Small world networks have two structural properties [48], the *small world effect* and *clustering*. The small world effect is the concept where any two nodes in a network are connected to each other through a small path. Quantitatively, this concept is captured using the metric called *average path length* (APL) which for any network, gives a value representing on average how far apart any pair of nodes lie in the network. Random networks also exhibit this property as typical APL values for any random network are low [32].

The second property of small world networks is *Clustering* which is also referred to as *transitivity* [26] or the *fraction of transitive triples* in a network [46]. This is to avoid confusion from the concept of *Community Structures* or *Clusters* [11,4]. The concept of clustering is the idea where two nodes having a common neighbour have a high tendency to be connected to each other. This property can be measured through the metric called clustering coefficient [48]. When compared to a random network, this property is only present in real world networks as in case of random networks, every node has an equal probability of connecting to any other node, thus transitive triples or triads¹ are rarely formed in a random network.

Many models have been proposed that artificially generate networks with small world properties. These models are quite useful as they can be used to construct artificial networks with desired properties and sizes that mimic real world networks. These generated networks can then serve as test beds to facilitate various experimental and empirical studies. Furthermore, these models provide us with insight into many real world systems and the methods to identify their structure and the methods to analyze them.

In this paper, we introduce a novel algorithm to generate small world networks from random networks. The classical model of Watts and Strogatz [48] suggests that, with the introduction of randomness in an ordered world, we can obtain small world networks. We demonstrate that with the introduction of small ordered components in a random world, we can still obtain a small world network. We then discuss the concept of community structures present in many real world networks and show how this model can be used to generate small world networks with community structures. Experimental results clearly show that the proposed model can be used to generate small world as well as clustered small world networks.

¹ The term was first used by sociologist Simmel [39] to represent the structure of social networks. It represents three people connected through three links forming a clique of size 3.

The novelty of our approach lies in the idea that we control the connectivity patterns of nodes in the generated small world network. By this, we mean that we embed structurally well defined groups of nodes (cliques or densely connected set of nodes) as compared to most of the other approaches present in the literature where the sole objective is to introduce triads to increase the overall clustering coefficient, thus generating small world graphs. Although, the real world social networks contain many triads [39,42], but this does not mean that they contain only triads. Cliques of various sizes must be present in artificially generated small world networks making them closer to real world networks. The proposed model allows this flexibility that cliques of different sizes can be introduced in a network using this model.

The proposed model also allows extension to generate small world graphs with clusters. To the best of our knowledge, all the other models of small world graphs do not generate clustered graphs. Since the presence of clusters has been widely studied in social and other networks [37], this model provides an interesting alternative to existing models where we want to generate clustered graphs with small world properties.

Another important property of real world networks observed by Albert and Barabasi [2] is the degree distribution of nodes following power-law also known as scale free distribution. Networks having power-law degree distribution are also known as scale free networks. In this paper we focus only on small world networks and their relationship with random networks and thus limit our discussion throughout this paper to the two structural properties of small world networks, the clustering coefficient and the average path length. We also limit our literature review to models that generate only small world networks without scale free properties although these models are briefly cited for the sake of completion.

The article is structured as follows: In the next section we review different models of small world networks. In section 3 we give the details of the proposed model to generate small world networks based on random networks. Next, we discuss the concept of community structures and how the proposed model can be extended to generate small world networks with community structures. Finally we conclude in section 5 giving future research directions.

2 Related Work

First we discuss the classical model of Watts and Strogatz (WS) [48]. The model starts with a completely regular network of n vertices where each vertex is connected to its k nearest neighbors, for a given value of k . This forms a regular graph as shown in Figure 1(a)². Then, each edge is rewired with a given probability p by choosing randomly a new vertex to connect. For a regular graph, the clustering coefficient is very high as neighbors are well connected to each other forming triads. The average path length is also very high as nodes only connect to their neighbors and distant nodes lie far apart from each other. Randomly rewiring a few edges introduces edges connecting nodes lying at long distances, and as a result, the overall average path length decreases generating a small world network as shown in Figure 1(b). Continuing the process of random rewiring, we end up rewiring all the edges in the initially generated regular graph obtaining a completely random network as shown shown in Figure 1(c).

² All the graphs in this article have been generated using Tulip <http://tulip.labri.fr/TulipDrupal/> which is an open source software for the analysis and visualization of large size graphs.

An important observation here is that we start with a completely regular network and the introduction of randomness transforms the graph to a small world network.

Although this basic model was published earlier, the study of these models started with another model published later by Watts as alpha model [47]. The idea is to capture two extreme worlds, a completely ordered world and a completely random world with a set of interaction rules. The completely ordered world is represented by what the authors call, caveman graphs. These graphs consists of isolated cliques called caves. Within each cave, everyone is connected to every other person but people from different caves do not communicate to each other. As a new person connects to someone from a cave, it almost immediately connects to everyone in the same cave. This results in the presence of high clustering coefficient as large size cliques also contain triads. At the other extreme, is the world where everyone can communicate to every other person irrespective of their previous connectivity. The probability of two people communicating is equal and thus new connections are formed completely at random. Based on these two interaction rules, we could construct a network evolving over time where new connections are alternately made according to one of the two specified interaction rules. Between the two extreme worlds, there lies a set of networks where order and randomness find an equilibrium such that there is enough order to have caves resulting in high clustering coefficient and there is enough randomness to have randomly connected nodes resulting in low average path length of the network. This set of networks has small world properties and can be tuned using the alpha parameter of how new connections chose the interaction rules. The idea is quite simple and tries to find a balance between an ordered graph and a random graph to generate small world networks.

Newman and Watts [29] proposed a slightly modified model to the classical model. In this model, instead of rewiring links between sites, as in the classical model of Watts and Strogatz, new links are added between pairs of sites chosen at random. No links are removed from the underlying regular network. This ensures that any region does not become disconnected. Moreover this model is easier to analyze and draw numerical results.

Newman *et al.* [30] also study a network generation model with arbitrary degree distributions which can be used to generate only small world networks. The idea is to generate affiliation networks similar to co-authorship³ [24] networks using random bipartite graphs.

The above idea is also used by Guillaume and Latapy [12] as they identify bipartite graph structure as a fundamental model of complex networks by giving real world examples. The model proposed can be used to generate small world and scale free graphs or just small world graphs depending on the degree distribution imposed on the bipartite graph. The two disjoint sets of a bipartite graph are called *bottom* and *top*. At each step, a new *top* node is added and its degree d is sampled from a prescribed distribution which can either be a uniform distribution or a power-law distribution. To generate the unipartite version from this bipartite graph, all the bottom nodes connected to a common top node are connected to each other forming cliques. The process is repeated for each top node to obtain the entire network. The process is illustrated in Figure 2 where (a) shows a bipartite graph and (b) shows the unipartite graph generated from it. As a result of this model, we obtain random small world networks but the bound-

³ A network of authors where two authors are connected to each other if they publish an artifact together.

aries between the cliques cannot be identified as cliques overlap due to the inherent bipartite structure in the network. These boundaries can be implanted in the bipartite graph initially generated but it will no longer remain randomly generated and would be a modified random bipartite graph. This model is quite useful to generate small world networks having inherent bipartite structure but our goal is to have a generic model which is not dependent on a bipartite structure. Moreover our objective is to generate a small world network which emerges from the random connectivity of small well connected group of nodes. This makes our proposed model easier to comprehend and analyze.

Another variation to the basic WS model was proposed by Kasturirangan [6] to generate small world networks. The author proposes the idea that the small world effect in networks does not occur due to random re-wiring of edges but due to multiple scales of network. These scales are formed due to the presence of high degree nodes which are responsible for reducing the overall average path length of a network. The model starts with a regular network just as the WS model and introduces high degree nodes by selecting randomly chosen vertices. Again this model uses a regular network and differs from our objective where we want to have well defined boundaries between small groups of nodes, to demonstrate that a certain amount of randomness can occur between groups of nodes, and still the network can exhibit small world properties. This is different from the classical WS model where introducing random edges does not force well defined boundaries between well connected nodes.

Another model based on similar justifications was proposed by Kleinberg [18] who used the notion of searching in these networks and how people are so efficient in finding these short paths. Developing a model based on how individuals do it in real world networks, the model will be more realistic and close to how networks are structured. The model again starts with a regular network and add random links to it, but the probability of a random link decreases with their distance as measured in the regular graph. Again this research is directed towards the understanding of how short paths occur in networks but the underlying network is kept regular. A good review of these small world models can be found in [25].

Barabasi and Albert [2] proposed a model to generate graphs with only scale free properties. Other researchers have proposed a number of artificial network generation models such as [15, 8, 20, 45, 10, 19, 5, 44, 14]. These models generate small world and scale free networks. Most of these models introduce triads to increase the clustering coefficient of the entire network and nodes connect using the preferential attachment rule to have a scale free degree distribution. These models are extensively reviewed in [49]. The author has also proposed another model to generate clustered small world and scale free graphs based on several social traits, readers can refer [49](chapter 4) for more details.

Models focussing domains such social networks [33] have also been studied where the goal is to develop more realistic models based on actual data instead of generalized structural properties. One notable class of such algorithms is the exponential random graphs [40] but these algorithms do not focus on small world properties and are not considered in this study.

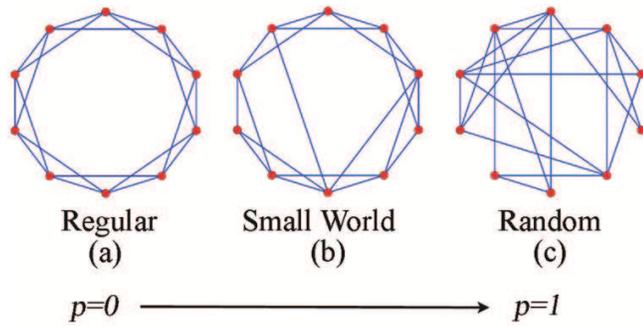


Fig. 1 Watts and Strogatz Model. From a Regular Network to a Random Network, where random rewiring of some edges in a regular network produces a small world network with high clustering coefficient and low average path length.

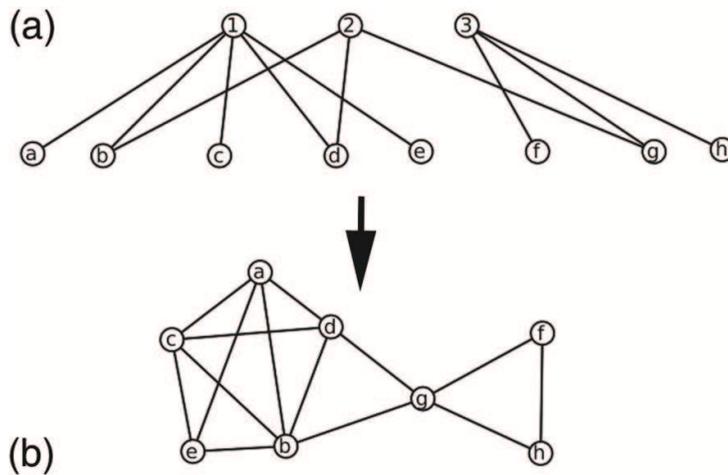


Fig. 2 Random Small World models from Bipartite Graphs.

3 Proposed Model

In this section we present the proposed model to generate small world networks from random networks. As opposed to WS model, we start with a random network. In the first step we generate a random network using the model of Erdos and Renyi[9]. In the next step, we replace each node in the random network with a triad, which is a set of three nodes connected through three edges forming a clique of size 3. In the final step, for each edge, in the random network an edge is placed between the two triads that replaced the original nodes in the initial random network connecting one of the randomly selected nodes from each triad. The final network thus obtained exhibits the two structural properties of a small world network.

Figure 3 illustrates the steps of the proposed model. In Figure 3(a) the random network generated using [9] is shown with 5 nodes. Figure 3(b) represents the second step of the proposed model where all the nodes are replaced with triads. Finally in step

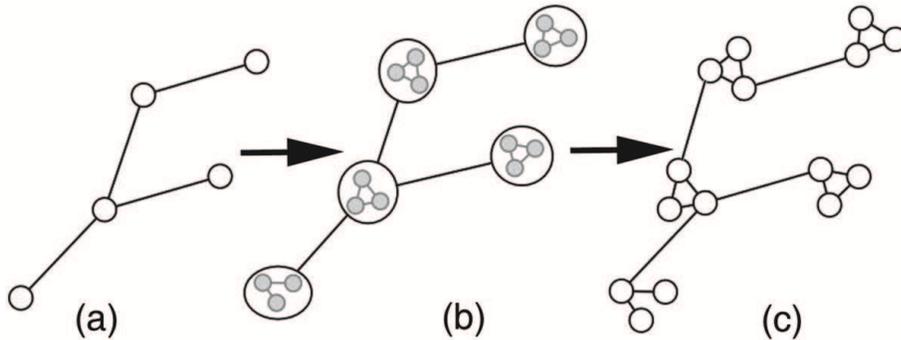


Fig. 3 Steps of the Proposed Model. (a) Random Network (b) Nodes replaced by triads (c) Edges connected to randomly selected newly added nodes.

3 shown in Figure 3(c), edges connecting the initial random network are connected to the one of the randomly selected nodes of the triads that replaced that particular node in step 2. Steps 2 and 3 transforms the random network into a small world network.

The introduction of triads increases the overall clustering coefficient of the network which is one of the basic properties of small world networks. The connectivity of these newly introduced triads in the network is based on the initially generated random network. We already know that random networks also have the property of small world effect [32] thus the average path length in the generated network remains low. It is easy to show that since the proposed model replaces nodes with triads, but the connectivity of these triads is governed by the initially generated random network, the average path length of the entire network remains low for the newly generated network.

Networks generated using this model indeed generates small world networks as shown from the values of clustering coefficient and average path length in Table 1. Since the metric clustering coefficient measures the presence of triads in a network, replacing triads for nodes in a random network statistically satisfies the properties of a small world network with high values of clustering coefficient but for real world networks, larger size cliques may also exist. A simple modification to this basic model is, instead of using triads, we can use cliques of larger sizes. Table 1 also shows the values of clustering coefficient and average path length when cliques of size 5 are used instead of 3. This value can be used as a parameter for the proposed model along with a randomly generated network which is given as input. Clearly the network thus generated also has small world properties. This idea can be further expanded to replace cliques of varying sizes to generate a small world network and need not to be a constant value. This can be a range between two constants or can be selected from an arbitrary distribution depending on how and why this artificially generated network is going to be used. Domain knowledge can also be used to determine the size of these cliques which can vary drastically from one domain to the other.

Experimental results are tabulated in Table 1 where we generate artificial networks for two different sizes of networks with different node-edge densities. This is to demonstrate the robustness of the proposed model. We use the ER model, the WS model and the proposed model for two different values cliques sizes 3 and 5. Results clearly show that the proposed model generates a small world network when compared to the

Table 1 Artificial Networks generated using Erdos and Renyi Random Graph Model, Watts and Strogatz Model and the Proposed Model for two different clique sizes (k) 3 and 5. APL = average path length, CC = Clustering Coefficient.

| Size | | Random Network | | Watts and Strogatz | | Zaidi $k = 3$ | | Zaidi $k = 5$ | |
|-------|-------|----------------|-------|--------------------|-------|---------------|-------|---------------|-------|
| Nodes | Edges | APL | CC | APL | CC | APL | CC | APL | CC |
| 100 | 250 | 3.01 | 0.062 | 3.56 | 0.299 | 3.03 | 0.363 | 3.81 | 0.694 |
| 100 | 400 | 2.43 | 0.080 | 3.06 | 0.511 | 2.48 | 0.306 | 2.53 | 0.402 |
| 1000 | 2500 | 4.48 | 0.003 | 5.70 | 0.331 | 4.45 | 0.393 | 6.40 | 0.696 |
| 1000 | 4000 | 3.55 | 0.007 | 4.47 | 0.418 | 3.55 | 0.352 | 3.71 | 0.367 |

statistics for a random network and the values are quite close to that generated by a WS Model.

Based on the random network initially generated, it is easy to calculate the number of nodes and edges the generated small world network will have using the proposed model. For clique size k and given a random network with nodes n and edges e , the proposed model generates a network with nodes n' given by equation 1 and edges e' given by equation 2.

$$n' = k * n \quad (1)$$

$$e' = e + \frac{k * (k - 1)}{2} * n \quad (2)$$

This simple experiment with four artificially generated graphs confirmed our hypothesis to generate small world networks from random graphs. We performed an extensive experiment using the proposed model where we generated 100 graphs each with $k=3$ and $k=5$ for different graphs sizes and densities. The APL and CC values are plotted for the generated graphs in Figure 4. For this experiment, we generated random graphs between sizes 100 and 1000 with edge/node ratio randomly generated between [2, 6]. Results clearly show that the proposed model for these parameters consistently generated small world networks as average APL values for the generated 100 graphs were 5.38 and 5.31, and the average CC values were 0.44 and 0.52 for $k=3$ and $k=5$ respectively.

4 Community Structures in Random and Small World networks

Another important property of networks is the presence of *community structures* or *clusters*. It is defined as the decomposition of nodes into ‘Natural Groups’ [36]. A more precise and mathematically quantifiable definition of a cluster is a set of vertices with high interconnectivity among vertices of the same cluster and low connectivity of vertices of different clusters [4]. There are several cluster evaluation metrics that try to evaluate the quality of a clustering algorithm on the basis of this ratio [50].

One such metric is the *Relative Density* [21] (RD). It calculates the ratio of the edge density inside a cluster to the sum of the edge densities inside and outside that cluster. The final Relative Density is the averaged sum of the these individual relative densities for all clusters and is given by the following equation:

$$RD = \frac{deg_{int}(C_i)}{deg_{int}(C_i) + deg_{ext}(C_i)} \quad (3)$$

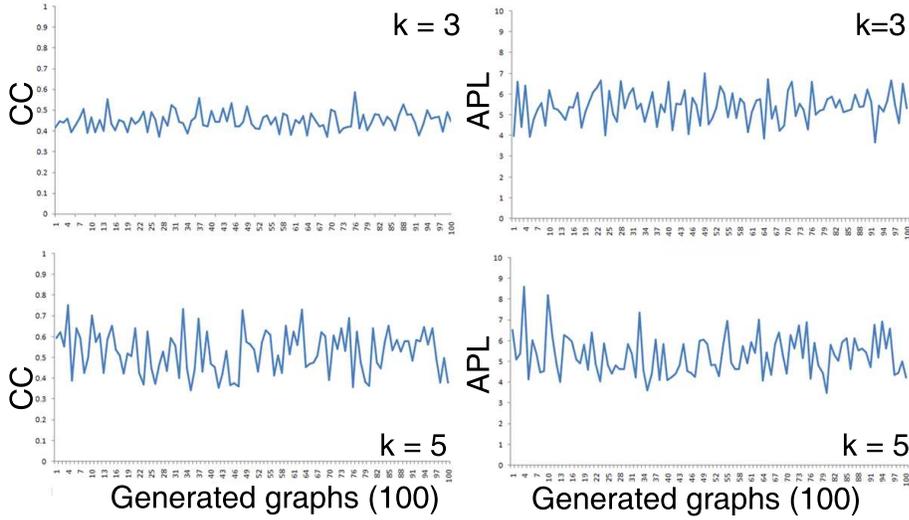


Fig. 4 APL and CC values calculated for randomly generated 100 small world graphs using the proposed model. For $k=3$ and $k=5$ over 100 graphs, the average CC values are 0.44 and 0.52, and average APL values are 5.38 and 5.31 respectively.

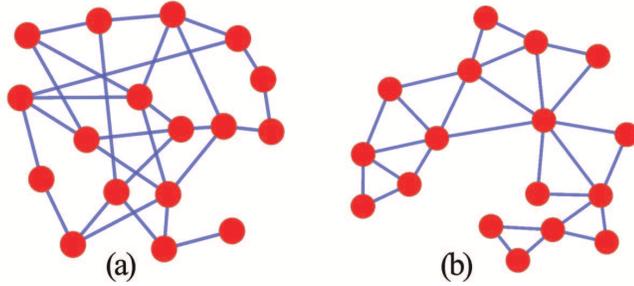


Fig. 5 Consider two graphs with same number of nodes and edges and thus having the same density in terms of number of nodes and number of edges. (a) Nodes well connected to each other forming quads, (b) Nodes sharing neighbors to form triads. Clustering Coefficient for graph (a) is 0.0 and (b) is 0.69 representing the absence of triads in graph (a). This example shows that nodes can be densely connected even if the clustering coefficient is low.

In the above equation, the term $deg_{int}(C_i)$ is the internal degree of cluster C_i defined as the number of edges connecting vertices in cluster C_i and the term $deg_{ext}(C_i)$ is the external degree of cluster C_i defined as the number of edges connecting vertices from cluster C_i to vertices of other clusters.

This definition of cluster is quite different from the metric earlier discussed called clustering coefficient. Clustering coefficient only tries to capture the presence of triads in a network where as a cluster tries to capture the notion of how densely connected a set of vertices is, regardless of the presence of triads. Two graphs are shown as a simple example in Figure 5 to highlight the differences of the two metrics. Both these graphs have 17 nodes and 28 edges making their node-edges density equal. The graph of

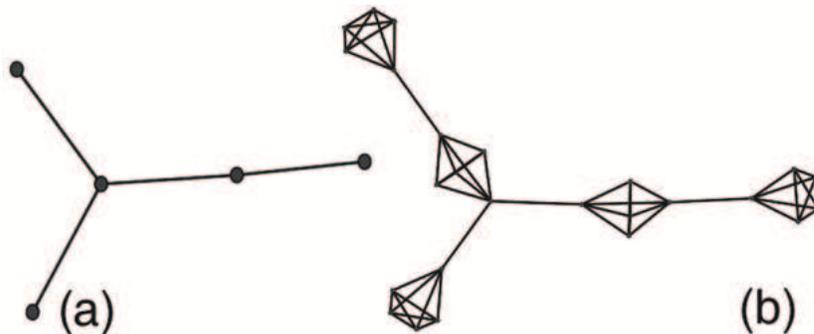


Fig. 6 (a) Shows a Random network with 5 nodes and 4 edges. (b) Shows the Small World network generated using the proposed model for $k=5$. We can consider each clique as a cluster in this generated small world network.

Figure 5(a) is built using quads, which we define as a set of 4 vertices connected through 4 edges forming a cycle (or a circuit). This clearly suggests that there are no triads in this graph. On the other hand, the graph of Figure 5(b) is built using triads. The clustering coefficient for the two graphs is 0.0 and 0.69 respectively. Consider if these two graphs are subgraphs of a another graph and we want to group them as clusters, the clustering co-efficient would only allow graph (b) to be clustered whereas in terms of the definition of a cluster, both have equal internal degree and either both should be clustered or both should not be clustered depending upon the overall topology of the network.

Coming back to the proposed model, lets consider another example of a random graph and the generated small world graph shown in Figure 6. Clique size k is set to 5 for this generation of small world graph. If we compare our definition of a cluster to this small world network, clearly we can see that the replaced cliques are extremely good candidates to be considered as clusters or natural groups in the network. The internal degree of each clique is the highest possible value since every node is connected to every other node in a clique. The external degree is not very high and in some cases, the external degree is as low as possible which is 1 for the cluster to be connected to the entire network.

Calculating the Relative Density (RD) from equation 3, we get the value of 0.86 which suggests a very good clustering in terms of RD. The idea can be further expanded to replace the nodes in the initial random network with not only cliques of larger sizes but with densely connected set of nodes. These densely connected group of nodes have certain social implications from our real world. People usually socialize in groups, as we have a group of friends at school, at work, in the family where every person knows the other person. This justifies the introduction of cliques or densely connected set of nodes in the graphs that are supposed to represent real world structures from our society. These closely knit group of nodes represent clusters of people in our real world. Thus the proposed model can be used to produce clustered small world artificial networks. An important perspective to this application of the proposed method is that the clusters connect to each other randomly and with minimal external degree.

This connectivity pattern is different from how Real world networks connect to each other as clusters are not so well isolated from the rest of the network. One solution to

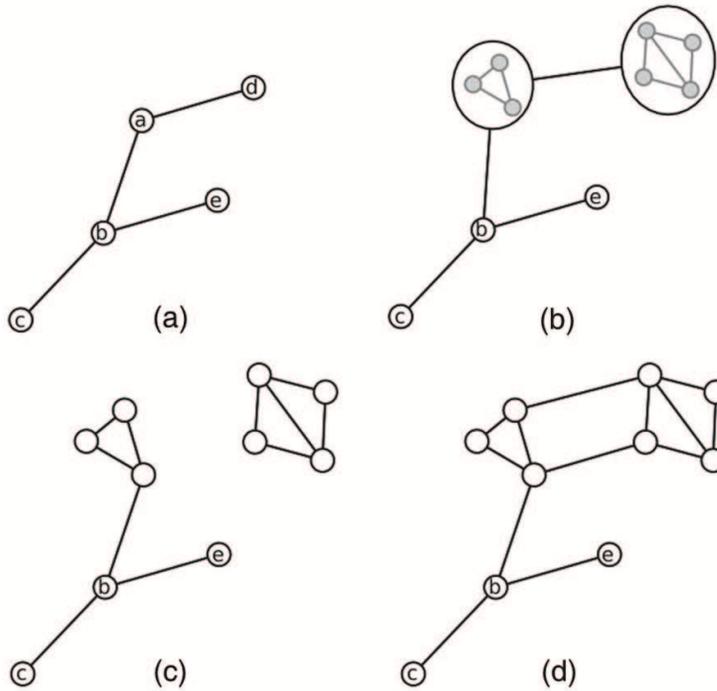


Fig. 7 (a)Random Network with 5 nodes and 4 edges.(b)Nodes a and d replaced by densely connected set of vertices.(c)Edge between the two newly introduced set of vertices is removed.(d)Multiple edges are introduced to connect the two newly introduced set of vertices.

resolve this interconnectivity of the clusters is that currently in the proposed model, we only replace nodes with cliques in the initially generated random network. We can also replace these single edges between clusters with more than one edges to increase the external degree of a cluster. This step will create fuzzy boundaries between clusters depending upon how high the external degree of a cluster is as compared to the internal degree. We summarize the modifications proposed to the basic models below:

- Instead of replacing nodes with cliques, densely connected set of vertices can be chosen. This requires four parameters, the minimum and maximum size for the set of vertices, the internal node-edge density and the clustering coefficient for the set of vertices.
- The edges in between nodes of the initially generated random network can be replaced with more than one edges. This connects nodes of the two set of vertices with more than one edges. The number of edges replaced determines the external node-edge density of the set of vertices representing clusters. This number can be taken as input from the user or can be generated randomly between 1 and the maximum number of edges possible between the two set of vertices.

After the above two modifications, the proposed model can be used to generate clustered artificial networks with small world properties where the internal and external densities are parametrized. The low external degree and high internal degree of

clusters ensures a well structured small world network with well defined clusters. The interconnectivity of clusters is based on the initially generated random network and these clusters communicate to each other randomly.

The process is graphically depicted in Figure 7. Again we start with a random network as shown in Figure 7(a). In the second step, nodes a and d are replaced with densely connected set of vertices where the size of each group is again chosen randomly. The resultant graph is shown in Figure 7(b). In step 3, the edge connecting nodes a and d in the initial random network is removed to be replaced by multiple edges connecting vertices of the two newly introduced sets of vertices. The number of edges to be added in between these newly added set of vertices can be given as parameter and the nodes to be connected from the two sets with these edges can be chosen randomly.

It is easy to show that the network thus generated again is a small world network as the nodes are replaced with set of vertices with high clustering coefficients. The inter-connectivity of these clusters is based on the initially generated random network so it also exhibits the small world effect.

We mention another metric called *Modularity* (Q) [28] to evaluate the quality of clusters. This metric has also been used by clustering algorithms [27] to obtain clusters from a network. The metric values range from 0 to 1 where 0 suggests poor clustering and 1 suggests the best clustering possible. Mathematically, for a specific division of a network into \mathcal{C} clusters, a symmetric matrix c can be defined as $\mathcal{C} \times \mathcal{C}$ whose element c_{ij} is the fraction of all edges in the network that link nodes in community i to nodes in community j . For this clustering, the Modularity(Q) can be defined as:

$$Q = Tr\ c - \|c^2\| \quad (4)$$

The term $Tr\ c$ refers to the trace of the matrix c which gives the fraction of edges in the network that connect vertices in the same community. The term $\|c^2\|$ is the squared sum of the elements of matrix c .

Modularity tries to capture the idea that if a particular clustering gives no more internal degree than would be expected by a random network, then clustering this set of vertices would give a poor value. If we look at the proposed model, this is exactly what we are trying to mimic when we introduce densely connected set of vertices in place of randomly connected nodes. The newly added set of vertices is expected to have a high density and thus high internal degree and low connectivity with nodes of other clusters, to have low external degree.

Table 2 shows the results of generating clustered small world networks using the most basic form of proposed model. We replace nodes with cliques of fixed sizes instead of randomly choosing the size, density and clustering coefficient of vertices. These cliques are clustered to form communities in the generated networks. Single edges between the nodes in the random network are not replaced with more than one edge. The results are shown for two different sets of graphs with different sizes and node-edge densities. The low average path lengths and high clustering coefficients clearly show that the generated networks have small world properties. We have used RD and Q metric to evaluate the quality of clusters produced using this proposed model and the obvious high values justify the grouping of nodes as clusters. These high values can be easily be tuned to desired values. For instance, to decrease these values, we can introduce more inter-cluster edges or we can reduce intra-cluster densities.

Analyzing the proposed model, it is interesting to observe that we have used the idea of caves from the alpha model [47] to generate a small world network. When we

Table 2 Clustered Small World Networks generated using the proposed model. The small world properties of generated networks can be observed with low APL = average path length and high CC = Clustering Coefficient values.

| Random Graph | | Clique Size k | Generated Graph | | APL | CC | Q | RD |
|--------------|-------|-----------------|-----------------|-------|------|------|------|------|
| Nodes | Edges | No. of Clusters | Nodes | Edges | | | | |
| 10 | 25 | 10 | 100 | 475 | 3.28 | 0.90 | 0.94 | 0.90 |
| 10 | 25 | 5 | 50 | 125 | 2.94 | 0.70 | 0.67 | 0.78 |
| 100 | 400 | 10 | 1000 | 4900 | 5.12 | 0.85 | 0.91 | 0.85 |
| 100 | 400 | 5 | 500 | 1400 | 4.54 | 0.59 | 0.71 | 0.56 |

generate a random network in the first step, it is to decide the number of caves our small world network will have and how these caves would be connected. When we extend this basic model to replace single edges with more than one edge, we modify the inter-connectivity of caves which can be used as a tunable parameter to control the external degree of a cluster. The size of each cave, the density and clustering coefficient are other parameters that can help users to generate tunable clustered small world networks. Thus, this model is useful to develop networks with desired structural as well as statistical properties for a small world network.

We performed a simple experiment to test how the proposed model behaves if we recursively apply it to the generated network. The idea was to explore the possibility of using the proposed model to generate hierarchies in small world networks as several researchers have shown the presence of hierarchical organization of real world networks, specially using divisive and agglomerative clustering algorithms [11,27,1]. Recursive application of the proposed model kept the clustering coefficient high but the average path length increased linearly as the number of nodes increased in the network. This is because recursively replacing nodes with cliques resulted in increase in the distances between the nodes in the newly generated networks. To reduce these distances, new random connections were added to reduce the overall average path length of the network. Since this addition lead to major changes in the proposed model, we leave further discussion as part of future work and limit this study to generating networks without hierarchical structures.

Again this simple experiment with four artificially generated graphs confirmed our hypothesis to generate clustered small world networks from random graphs. We performed an extensive experiment using the proposed model where we generated 200 graphs where the parameters were changed after the first 100 graphs generated. The five parameters used are the number of cliques (or clusters) to be added in the graph. This is essentially the number of nodes in the initially generated random graph. Next two parameters are the minimum and maximum sizes of the cliques that will replace the nodes in the random graphs. Fourth parameter is the percentage of edges to be removed from the graph containing cliques which reduces the overall density of clusters. And finally the percentage of edges to be added between different clusters to increase the intercluster density of the clustered graph. For this experiment, we generated random graphs between sizes 100 and 1000. The clique sizes varied between [5,25] and were randomly generated for each cluster and for each graph. Both these ranges remain unchanged for all the 200 generated graphs. The percentage of edges to be removed from the cliques varied between [0,10] and [0,40] for the two sets of 100 graphs. Similarly the percentage of edges to be added between clusters varied between [0,10]

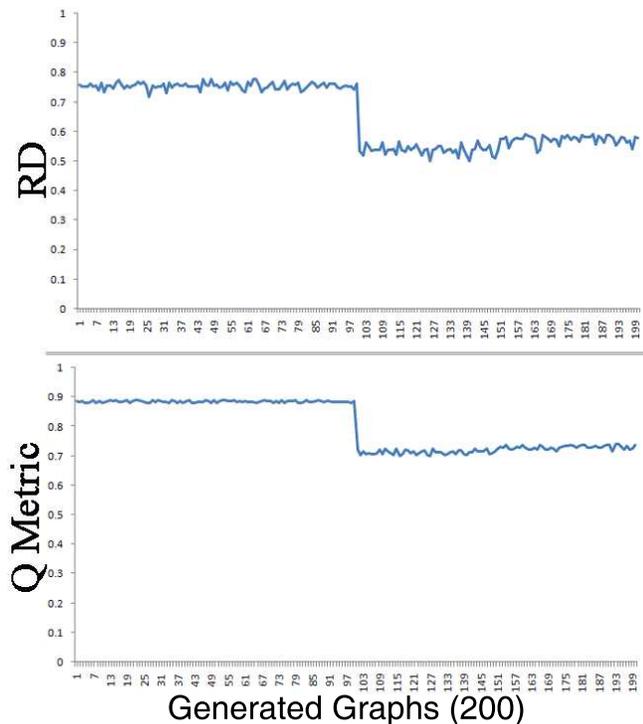


Fig. 8 RD and Q metric values calculated for randomly generated 200 clustered small world graphs using the proposed model. ‘Intercluster Edges’ and ‘Density of Cliques’ Parameters are changed after 100 clustered graphs.

and $[0,40]$. The percentages were calculated on the basis of the total number of edges present in the graph after the cliques were added to the graph. Once these clustered graphs were generated, the RD and Q values were calculated and consistently high values were obtained for the 200 graphs generated. These high values suggest that the generated graphs had indeed community structure clearly present. Results are shown in Figure 8.

One such graph is shown in Figure 9 where we generated a random graph of 10 nodes and 25 edges. These 10 nodes were then replaced by cliques of different sizes and 10% of the total edges of these cliques were removed. Thus making a graph of 67 nodes and 197 edges. The average clustering coefficient of this graph is 0.68 and the average path length is 3.29. Considering these clusters, we calculate the Q metric which turns out to be 0.86 and the RD value to be 0.76. This simple example demonstrates that indeed our model works well to generate small world clustered networks.

Considering the two extremes for our model, when we simply replace the nodes in a random graph with cliques representing communities, we have a small world network with high clustering coefficient and small average path length. As the model can be parameterized, if we keep reducing the number of edges inside a cluster (intracluster edges) and keep increasing the number of edges between these clusters, we will end up with a completely random graph since all the intracluster edges will be removed (thus

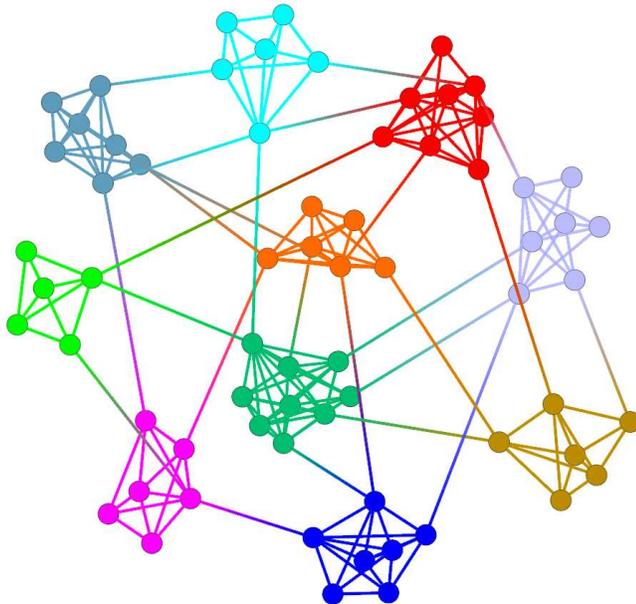


Fig. 9 Shows a graph generated using the proposed model where the initial random graph contained 10 nodes and 25 edges. These nodes were replaced by cliques of size 5-10. Color encoding represents nodes belonging to the same cluster with 25 intercluster edges.

removing all the triads) and many intercluster edges that randomly connect nodes from different clusters will increase. We performed this experiment to show how the small world graph changes to a random graph if the process of removing intracluster edges and adding intercluster edges is performed iteratively. The results are tabulated in Table 3 and Table 4. These results were generated when the initial random graph contained 100 nodes and 400 edges. These nodes were replaced by cliques of varying sizes between 5 and 10.

Table 3 shows the results when different percentages of the intracluster edges are removed from the clusters. We used the percentage ranges from 0% to 90% to remove edges from the cliques where 0% means no edge was removed from the cliques and 90% means 90% of the total possible edges inside clusters were removed. From Table?? and Figure10(a,c), we can clearly see that removing these edges linearly reduces the clustering coefficient of the generated graph while slightly increasing the average path length of the entire graph.

Table 4 shows the results when the intercluster edges are gradually added to the graph. We increased the edges by multiplying the originally present edges (for these generated graphs, we used $n=100$ and $e=400$) with a factor ranging between 1 and 10 where 1 means no change in the existing number of edges and 10 means original edges were replaced with $400 * 10=4000$ edges in this particular case. From Table?? and Figure10(b,d), we can clearly see that adding these edges linearly reduces the clustering coefficient of the generated graph while slightly decreasing the average path length of the entire graph.

Table 3 Shows how the values of Clustering Coefficient (CC), Average Path Length (APL), Q and Relative Density (RD) behave as we remove intracluster edges reducing the number of edges present inside clusters. The values are plotted for removal percentage where for example 90% signifies that 90% of the total edges present inside a network were remove.

| Nodes | Edges | APL | CC | Q | RD | Removed Edges | Added Edges |
|-------|-------|------|------|-----|-----|---------------|-------------|
| 707 | 2637 | 4.99 | .73 | .84 | .72 | 0 | 0 |
| 701 | 2406 | 4.92 | .62 | .82 | .68 | 10% | 0 |
| 687 | 2108 | 5.1 | .52 | .79 | .65 | 20% | 0 |
| 705 | 1991 | 5.3 | .44 | .78 | .62 | 30% | 0 |
| 665 | 1615 | 5.3 | .32 | .73 | .53 | 40% | 0 |
| 694 | 1505 | 5.6 | .25 | .71 | .48 | 50% | 0 |
| 734 | 1392 | 5.9 | .18 | .69 | .40 | 60% | 0 |
| 683 | 1050 | 5.7 | .09 | .59 | .25 | 70% | 0 |
| 736 | 911 | 6.38 | .05 | .53 | .19 | 80% | 0 |
| 662 | 620 | 5.7 | .007 | .30 | .09 | 90% | 0 |

Table 4 Shows how the values of Clustering Coefficient (CC), Average Path Length (APL), Q and Relative Density (RD) behave as we add intercluster edges increasing the number of edges present between clusters. The originally existing edges are multiplied by the number of times for example 9 signifies that (400 * 9) intercluster edges were added to the network.

| Nodes | Edges | APL | CC | Q | RD | Removed Edges | Added Edges |
|-------|-------|------|-----|-----|-----|---------------|-------------|
| 707 | 2637 | 4.99 | .73 | .84 | .72 | 0 | 1 |
| 706 | 3038 | 3.8 | .56 | .73 | .56 | 0 | 2 |
| 700 | 3403 | 3.4 | .43 | .64 | .46 | 0 | 3 |
| 684 | 4193 | 3.05 | .29 | .51 | .34 | 0 | 4 |
| 659 | 4342 | 2.8 | .21 | .43 | .27 | 0 | 5 |
| 695 | 4959 | 2.8 | .20 | .42 | .26 | 0 | 6 |
| 701 | 5380 | 2.7 | .18 | .39 | .24 | 0 | 7 |
| 696 | 5753 | 2.6 | .15 | .36 | .22 | 0 | 8 |
| 692 | 6117 | 2.6 | .13 | .33 | .20 | 0 | 9 |
| 707 | 6623 | 2.6 | .13 | .32 | .19 | 0 | 10 |

From this final set of experiments we can observe the linear change in the Q and RD values that decreases with the decrease in the intracluster edges and increase in the intercluster edges. These values along with the low values of clustering coefficient represent the absence of triads and community structures in a graph where nodes connect randomly to each other.

5 Conclusion and Future Research Perspective

In this paper, we have introduced a novel algorithm to generate small world networks based on random networks. The novelty of the proposed algorithm is the introduction of inherent randomness in the connectivity patterns of groups of nodes. We demonstrate that with the introduction of some order in a random world, we can still produce graphs with small world properties. The model is further extended to generate small world networks with well defined community structures. An extended literature review suggests that this is the first attempt to generate clustered small world graphs. Experiments clearly show that the generated networks exhibit desired small world and clustered small world properties.

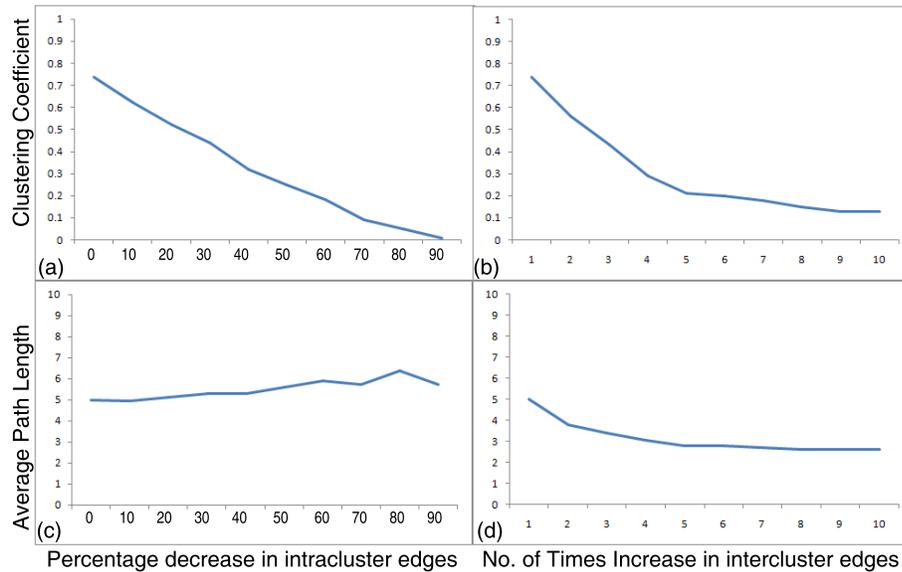


Fig. 10 Shows the behavior of Clustering Coefficient and Average Path length as we decrease intracluster edges and increase intercluster edges. (a) Clustering coefficient drops linearly as intracluster edges are removed. (b) Clustering Coefficient again drops as intercluster edges are increased. (c) Average path length increases slightly as intracluster edges are removed. (d) Average path length decreases slightly as intercluster edges are added to the network.

We also tested the proposed model to generate hierarchical structures in small world networks. We recursively used the model to generate hierarchies which resulted in networks with high clustering coefficient but the average path length also increased linearly with the introduction of each level. We intend to explore this direction to be able to generate small world networks with hierarchies. We also intend to extend this study for scale free networks where the idea is to have clustered scale free networks generated artificially.

References

1. D. Auber, Y. Chiricota, F. Jourdan, and G. Melançon. Multiscale visualization of small world networks. In *INFOVIS '03: Proceedings of the IEEE Symposium on Information Visualization*, pages 75–81, 2003.
2. A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
3. S. Bilke and C. Peterson. Topological properties of citation and metabolic networks. *Rev. E*, 64:036106, 2001.
4. U. Brandes and T. Erlebach. *Network Analysis : Methodological Foundations (Lecture Notes in Computer Science)*. Springer, March 2005.
5. M. Catanzaro, G. Caldarelli, and L. Pietronero. Assortative model for social networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 70(3):1–4, 2004.
6. R. K. Department and R. Kasturirangan. Multiple scales in small-world networks. In *Brain and Cognitive Science Department, MIT*, 1999.
7. C. Discazeaux, C. Rozenblat, P.-Y. Koenig, and G. Melançon. Territorial and topological levels in worldwide air transport network. In *15TH European Colloquium on Theoretical and Quantitative Geography, Montreux, Switzerland*, 2007.

8. S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advances in Physics*, 51:1079–1187, June 2002.
9. P. Erdős and A. Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.
10. P. Fu and K. Liao. An evolving scale-free network with large clustering coefficient. In *ICARCV*, pages 1–4. IEEE, 2006.
11. M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99:8271–8276, 2002.
12. J.-L. Guillaume and M. Latapy. Bipartite graphs as models of complex networks. In *Workshop on Combinatorial and Algorithmic Aspects of Networking (CAAN)*, LNCS, volume 1, 2004.
13. R. Guimera, S. Mossa, A. Turttschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(22):7794–7799, 2005.
14. W. Guo and S. B. Kraines. A random network generator with finely tunable clustering coefficient for small-world social networks. In *CASON '09: Proceedings of the 2009 International Conference on Computational Aspects of Social Networks*, pages 10–17, Washington, DC, USA, 2009. IEEE Computer Society.
15. P. Holme and B. J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65:026107, 2002.
16. H. Jeong, B. Tomber, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.
17. A. Kirman. The economy as an evolving network. *Journal of Evolutionary Economics*, Springer, 7(4):339–353, 1997.
18. J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, Aug. 2000.
19. K. Klemm and V. M. Eguiluz. Growing scale-free networks with small world behavior. *Physical Review E*, 65:057102, 2002.
20. J.-G. Liu, Y.-Z. Dang, and Z. tuo Wang. Multistage random growing small-world networks with power-law degree distribution. *Chinese Phys. Lett.*, 23(3):746, Oct. 31 2005. Comment: 3 figures, 4 pages.
21. M. Mihail, C. Gkantsidis, A. Saberi, and E. Zegura. On the semantics of internet topologies, tech. rep. gitcc0207. Technical report, College of Computing, Georgia Institute of Technology, Atlanta, GA, USA, 2002.
22. J. M. Montoya and R. V. Solé. Small world patterns in food webs. *Journal of Theor. Biol.*, 214:405–412, February 2002.
23. C. Moore and M. E. J. Newman. Epidemics and percolation in small-world networks. *Phys. Rev. E*, 61:5678–5682, 1999.
24. M. E. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Phys Rev E Stat Nonlin Soft Matter Phys*, 64(1 Pt 2), July 2001.
25. M. E. J. Newman. Models of the small world. *Journal of Statistical Physics*, 101:819–841, 2000. 10.1023/A:1026485807148.
26. M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167, 2003.
27. M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69:066133, 2004.
28. M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69(2 Pt 2):026113, February 2004.
29. M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263:341–346.
30. M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 1):2566–2572, February 2002.
31. R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scalefree networks. *Phys. Rev. Lett.*, 86(14):3200–3203, 2001.
32. A. Reka and Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, June 2002.
33. G. Robins, P. Pattison, Y. Kalish, and D. Lusher. An introduction to exponential random graph (p) models for social networks. *Social Networks*, 29(2):173–191, May 2007.
34. D. Rosen, G. A. Barnett, and J.-H. Kim. Social networks and online environments: when science and practice co-evolve. *Social Netw. Analys. Mining*, 1(1):27–42, 2011.

-
35. C. Rozenblat, G. Melançon, and P.-Y. Koenig. Continental integration in multilevel approach of world air transportation (2000-2004). *Networks and Spatial Economics*, 2008.
 36. S. E. Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, August 2007.
 37. J. Scott. Social network analysis: developments, advances, and prospects. *Social Network Analysis and Mining*, 1:21–26, 2011. 10.1007/s13278-010-0012-6.
 38. J. P. Scott. *Social Network Analysis: A Handbook*. SAGE Publications, January 2000.
 39. G. Simmel and K. H. Wolff. *The sociology of Georg Simmel / translated and edited with an introduction by Kurt H. Wolff*. Free Press, Glencoe Ill, 1950.
 40. T. A. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, Dec. 2006.
 41. R. Solis. Visualizing stock-mutual fund relationships through social network analysis. *Global Journal of Finance and Banking Issues*, 3(3), 2009.
 42. C. E. Tsourakakis, P. Drineas, E. Michelakis, I. Koutis, and C. Faloutsos. Spectral counting of triangles via element-wise sparsification and triangle-based link recommendation. *Social Netw. Analys. Mining*, 1(2):75–81, 2011.
 43. A. Wagner and D. Fell. The small world inside large metabolic networks, Aug. 21 2000.
 44. J. Wang and L. Rong. Evolving small-world networks based on the modified ba model. *Computer Science and Information Technology, International Conference on*, 0:143–146, 2008.
 45. L. Wang, F. Du, H. P. Dai, and Y. X. Sun. Random pseudofractal scale-free networks with small-world effect. *The European Physical Journal B - Condensed Matter and Complex Systems*, 53:361–366, 2006.
 46. S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.
 47. D. Watts. Networks, dynamics and the small-world phenomenon. *American Journal of Sociology*, 105 (2):493–527, 1999.
 48. D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.
 49. F. Zaidi. *Analysis, Structure and Organization of Complex Networks*. PhD thesis, Ecole doctorale de Mathématiques et Informatique, Université de Bordeaux I, France, 2010.
 50. F. Zaidi, D. Archambault, and G. Melançon. Evaluating the quality of clustering algorithms using cluster path lengths. In *Advances in Data Mining. Applications and Theoretical Aspects, 10th Industrial Conference, ICDM*, pages 42–56, 2010.