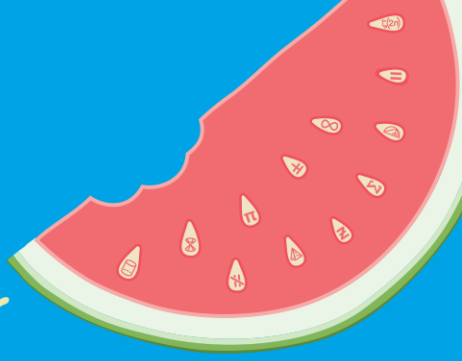
AMSI VACATIONRESEARCH SCHOLARSHIPS 2021–22

Get a taste for Research this Summer



Properties of Random Networks

David Chen
Supervised by Dr Nathan Ross
University of Melbourne

February 28, 2022

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Contents

Abstract

Many existing random network models such as Erdős–Rényi and preferential attachment have deficiencies in modelling real-world networks. In particular, real-world networks are often highly clustered while those models fail to exhibit that property. This paper introduces a new general embellishment of network models that rectifies some of those deficiencies, through the random addition of edges between vertices that have common connections. Some general results are also presented and proved.

1 Introduction

1.1 Background and Introduction

Graph structures present in many real-world systems, such as friendship networks, the Internet and biological networks In particular, those networks are usually enormous with intrinsic randomness, making it difficult to study their structures on a global scale [1]. Therefore, random networks are natural candidates in modelling such systems. Due to the complexity of real-world network systems, it is only feasible to extract statistics out of empirical data. Some of the common statistics include:

- The degree distribution of the graph. This is the proportion of nodes in the graph with a given degree k. Many real-world graphs are observed to be approximately scale-free, meaning that the tail degree distribution is asymptotically proportional to $k^{-\gamma}$.
- The typical distance of the graph. This refers to the length of a path between two uniformly randomly chosen vertices. Real-world graphs are typically small-worlds (even ultra-small worlds). That is to say the typical distance is of order $O(log(n))$ with high probability.
- Components structures. The components here refers to the connected components, which are sets of nodes that can reach one another through the edges. Empirical data suggests real-world networks are usually highly-connected in the sense that there is a giant component almost the size of the network with a few other smaller ones.s
- Clustering coefficients. The clustering coefficient is the proportion of wedges (or v-shapes) in the graph with the closing edge present. Real-world networks are typically highly-clustered with their clustering coefficient being positive (not close to 0).

We will introduce the background knowledge and discuss those properties with more detail in section 2.

The general aim of random network models is to explain and model the properties of real-world networks through probabilistic or stochastic methods, as well as their structure as they grow. Many random network models have been developed over the years. The most well-studied ones include:

- Erdős–Rényi $\mathrm{ER}\big(n, \frac{\lambda}{n}\big)$
- Configuration Model $\mathrm{CM}(n, \mathbf{d})$
- Preferential Attachment $\mathrm{PA}_n^{(m,\delta)}(b)$

In section 3, we will introduce those models in more detail.

As summarised in table 1.1, those models are not highly clustered. The reason behind it is their tree-like local limit, which means they have a high number of uncompleted triangles. We will formalise the idea of local limits in section 4.

Random Network Model	Scale-free?	Small-world?	Highly-connected?	Highly-clustered
Erdős–Rényi	No	Yes	Yes	No
Configuration Model	Yes	Yes	Yes	No
Preferential Attachment	Yes	Yes	Yes	No

Table 1.1: A comparison of random network models.

We observe in real-world friendship networks, individuals are quite likely to become friends with their friends-of-friends, which motivates this research. Based on this observation, we introduce an embellishment by adding those missing friends-of-friends connections independently with some probability q. We will show that the method is applicable to a wide variety of random network model and improves their clustering behaviour, making the models highly-clustered. In section 5 we will introduce the embellishment in greater detail and derive a few other properties of embellished graphs.

1.2 Statement of Authorship

David Chen formalised the definitions of the embellishment, developed and proved properties of the embellishment.

Dr Nathan Ross developed the idea of the embellishment, supervised the project, assisted with proofs of some of the theorems and proofread this report.

2 Random Graphs and Their Properties

In this section we will introduce definitions used in the future sections. In subsection 2.1 we will formally define random graphs and related concepts. After that we will explore properties of real-world graphs and formulate them with the language of random graphs in subsection 2.2.

2.1 Graphs and Random Graphs

Definition 2.1 (Graphs)**.** *A* ***graph*** G *is a pair* $(V(G), E(G))$ *where* $V(G)$ *is the countable* ***vertex set*** *or* ***node set*** *and* $E(G)$ *the* ***edge set****. G is a* ***directed graph*** *if for all* $u, v \in V(G)$*,* u, v *are* ***adjacent*** *if and only if* $(u, v) \in E(G)$*. G is an* ***undirected graph*** *if for all* $u, v \in V(G)$*,* u, v *are* ***adjacent*** *if and only if* $\{u, v\} \in E(G)$*.*

For our study we will consider undirected graphs only, so unless otherwise specified all graphs mentioned in this paper are undirected. The next concept is the degree of a node, which is the number of edges attached to that node.

Definition 2.2 (Degree)**.** *Let G be a graph. The degree of a node $v \in V(G)$ is $d_G(v)$, the number of vertices adjacent to it. That is, $d_G(v) = |\{u \in V(G) | \{u, v\} \in E(G)\}|$.*

We will also use $D_G = d_G(V)$ to denote the degree of a uniformly randomly selected node V, regardless of whether the graph G is random. This random variable is also known as the typical degree of G. Now we will introduce paths and distances in graphs.

Definition 2.3 (Paths and Length of Paths)**.** *Let G be a graph and $u, v \in V(G)$. A* ***path*** *between u and v is a finite sequence of adjacent nodes $\mathbf{w} = (w_0, \ldots, w_k)$ such that $w_0 = u$, $w_k = v$ and $\{w_i, w_{i+1}\} \in E(G)$. The* ***length*** *of this path $\mathbf{w}$ is k.*

Note that paths in general can visit a node multiple times and form cycles and they do not need to be the most "direct" way to connect nodes. This concept leads to connectedness and graph distances.

Definition 2.4 (Connectedness and Connected Components)**.** *Let G be a graph and $u, v \in V(G)$. We say that u is* ***connected*** *to v if there is a path between u and v. The* ***connected component*** *of u is $\mathcal{C}(u)$, the set of nodes connected to u. The* ***giant component*** *of G is the largest connected component in G, denoted $\mathcal{C}_{max}$.*

Definition 2.5 (Graph Distance)**.** *Let G be a graph and $u, v \in V(G)$ connected. The* ***graph distance*** *between u and v is the length of the shortest path between u and v.*

Next, we define graphs with a fixed root node, which is useful in local convergence as we will see in section 4.

Definition 2.6 (Rooted Graph)**.** *A* ***rooted graph*** *is a pair (G, o) where G is a graph and $o \in V(G)$ a fixed vertex. Sometimes we simply write G when there is no ambiguity.*

We would expect two graphs G and H to be equivalent if we merely reordered the nodes since the structure of the graph remain unchanged. This leads to the definition of graph isomorphisms.

Definition 2.7 (Graph Isomorphism)**.** *Let G and H be graphs. We say G is* ***isomorphic*** *to H and write $G \cong H$ if there is a bijection $\phi : V(G) \to V(H)$ such that $\{v, w\} \in E(G)$ if and only if $\{\phi(v), \phi(w)\} \in E(G)$.*

Definition 2.8 (Rooted Graph Isomorphism)**.** *Let (G, o) and (H, v) be rooted graphs. We say (G, u) is **isomorphic** to (H, v) and write $(G, o) \cong (H, v)$ if there is a bijection $\phi : V(G) \to V(H)$ such that $\phi(o) = v$ and $\{v, w\} \in E(G)$ if and only if $\{\phi(v), \phi(w)\} \in E(G)$.*

A random graph G is a graph structure with a fixed vertex set $V(G)$ yet a random edge set $E(G)$. Formally, we define them as a random vector of indicators denoting whether an edge is present in the graph.

Definition 2.9 (Random Graphs)**.** *A **random graph** G is a random vector $(\mathbb{1}_{\{u,v\}})_{u,v \in V(G)}$ where $V(G)$ is a determined set.*

We also use $(G_n)_{n>0}$ to denote a sequence of random graphs with increasing vertex set size $|V(G_n)|$ and study their properties as $n \to \infty$. This is because we are interested in how those models evolve as they grow in size and $n \to \infty$ is consistent with the nature that real-world networks are large. We will provide some examples of such properties in the next subsection.

Also, note that it is common practice to simply write G, $E(G)$ for random graphs and the random edges. We will adopt this convention in the future sections.

2.2 Properties of Random Networks

Properties of real-world random networks include scale-free, highly connected, highly clustered, small-world. We will define the above properties in this subsection. Note that there are also other properties of interest, such as assortativity (degree-degree dependencies), centrality (a measure of vertex importance) and community structures (highly-connected local subgraphs), which we will not go into great detail here.

To define scale-free, we need to define sparse first.

Definition 2.10 (Sparse)**.** *([1, p11] Let $(G_n)_{n>0}$ be a sequence of graphs. We call $(G_n)_{n>0}$ sparse if the empirical degree sequence converges to some non-negative integer distribution p_k,*

$$\left(\frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} \mathbb{1}_{\{d_{G_n}(v)=k\}} \right)_{k \geq 0} \to (p_k)_{k \geq 0} \, .$$

When G_n is random we can modify the above definition to make it a convergence in distribution or in probability.

Definition 2.11 (Scale-free)**.** *Let $(G_n)_{n>0}$ be a random graph sequence. Then we say G_n is **scale-free** if as $k \to \infty$,*

$$\sum_{i>k} p_i \sim k^{-r} \, .$$

Note that the definition of scale-free can be relaxed with slowly-varying functions [1, p11-p12], which we will not discuss in detail here.

Definition 2.12 (Highly Connected)**.** *([1, p14]) Let $(G_n)_{n>0}$ be a sequence of random graphs. We say that G_n is **highly-connected** if the size of the giant component $\mathcal{C}_{max}$ is $\Theta(n)$.*

Usually the giant component is unique, i.e. the size of the second largest component is $o(1)$.

The clustering coefficient of a graph comes with a few variants. We will use the local clustering coefficient here since it is relatively easy to compute. This clustering coefficient measures the average number of ordered triangles with a vertex at v over the number of ordered wedges with their turning points at v.

Definition 2.13 (Local Clustering Coefficient)**.** *([1, p17]) Let G be a graph (random or not random). Define the number of triangles through a node v*

$$\Delta_G(v) = \sum_{u,w \in V(G)} \mathbb{1}_{\{u,v\} \in E(G)} \mathbb{1}_{\{v,w\} \in E(G)} \mathbb{1}_{\{w,u\} \in E(G)}$$

and the clustering coefficient of a specific node v

$$\mathrm{CC}_G(v) = \frac{\Delta_G(v)}{d_G(v)(d_G(v)-1)}\,.$$

*Then the **local clustering coefficient** of G is*

$$\overline{\mathrm{CC}}_G = \frac{1}{|V(G)|} \sum_{v \in V(G)} \mathrm{CC}_G(v) = \frac{1}{|V(G)|} \sum_{v \in V(G)} \frac{\Delta_G}{d_G(v)(d_G(v)-1)}\,.$$

The highly clustered property of real-world networks reveals the fact that wedges are often complete. In the example of friendship networks, it means an individual's friends are likely to be friends with each other as illustrated in figure 2.1.

Figure 2.1: Your friends are friends (left) vs your friends are not friends (right). In real-world networks, (left) is common.

Definition 2.14 (Highly Clustered)**.** *A sequence of graphs (not random) are said to be **highly clustered** if*

$$\liminf_{n \to \infty} \overline{\mathrm{CC}}_{G_n} > 0\,.$$

When the graph sequence $(G_n)_{n>0}$ is random, (CC_n) is a sequence of random variables, which makes the concept of highly clustered tricky to define. We provide a weak definition and a much stronger one below.

Definition 2.15 (Weakly Highly Clustered)**.** *A sequence of random graphs is **weakly highly clustered** if*

$$\lim_{n \to \infty} \mathbb{E}\big(\overline{\mathrm{CC}}_{G_n}\big) > 0\,.$$

Definition 2.16 (Strongly Highly Clustered)**.** *A sequence of random graphs is **strongly highly clustered** if for some constant $c > 0$,*

$$\overline{\mathrm{CC}}_{G_n} \xrightarrow{\mathbb{P}} c\,.$$

The weak version shown in definition only requires the expectation of the clustering coefficient to converge to a positive constant, whereas the strong version presented in definition requires the random variable to converge in probability, which implies the weak version. Note that a convergence in distribution of $\overline{\mathrm{CC}}_{G_n}$ is sufficient to meet the definition of the strong version since the limiting random variable is a constant.

Finally we will define the small-world property.

Definition 2.17 (Typical Distance)**.** *Let G be a random graph. The* ***typical distance*** *$L(G)$ is the graph distance of any two uniformly randomly chosen vertices, conditioned on them being on the same connected component.*

Definition 2.18 (Small-world)**.** *([1, p15]) Let $(G_n)_{n>0}$ be a sequence of random graphs. G_n is said to be a* ***small-world*** *if there exist a constant $K > 0$ such that*

$$\mathbb{P}(L(G_n) \leq K \log n)) \to 1\,.$$

G_n is said to be an ***ultra-small world*** *if for every constant $\epsilon > 0$,*

$$\mathbb{P}(L(G_n) \leq \epsilon \log n)) \to 1\,.$$

3 Random Network Models

In this section we will introduce some of the most popular random network models, including Erdős–Rényi $\mathrm{ER}(n, \frac{\lambda}{n})$, Configuration Model $\mathrm{CM}(n, \mathbf{d})$ and Preferential Attachment $\mathrm{PA}_n^{(m,\delta)}(b)$.

3.1 Erdős–Rényi Models

The Erdős–Rényi random graph $\mathrm{ER}\big(n, p_n = \frac{\lambda}{n}\big)$ takes two parameters where n is the number of vertices in the graph and p_n is the probability any edge is present in the graph. In addition, we assume the edges between any pair of nodes are added mutually independently. The Erdős–Rényi random graph is one of the simplest random network models, yet it also displays many interesting features. Figure 3.1 shows an example of the Erdős–Rényi graph.

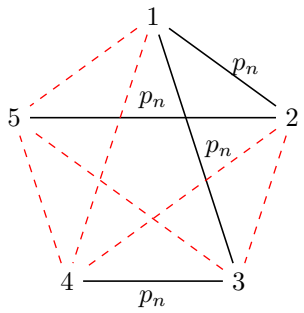


Figure 3.1: A realisation of the Erdős–Rényi model $\mathrm{ER}(n, p_n = \frac{\lambda}{n})$ with $\lambda = 2.5$ and $n = 5$. Black edges are present in the graph while the red edges are not.

3.2 Configuration Models

The configuration model builds the network in a reversed order. Starting with n nodes and a fixed degree sequence $\mathbf{d_n}$ for the graph, we try to pairs up the "stubs" randomly, as illustrated by figure 3.2. Note that self-loops and multi-edges may be generated in the process.

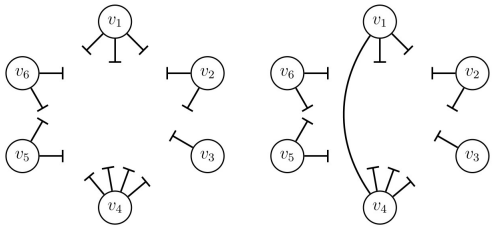


Figure 3.2: Illusting the configuration model [1, p216]

The ability to specify the degree distribution allows the configuration model to be scale-free. However, as we will see in Theorem 4.5, the configuration model cannot be highly-clustered.

3.3 Preferential Attachment

In the preferential attachment model, nodes are added one at a time to the graph and connected to the existing nodes randomly, with more weights placed on nodes with high degree.

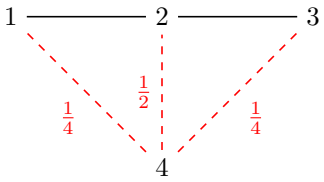


Figure 3.3: Preferential attachment: Vertex 2 is a high-degree vertex and so new vertices are more likely to be connected with it.

It can be shown that Preferential Attachment model are scale-free. Nevertheless, similar to the configuration model, its local limit is tree-like as shown by Theorem 4.6.

4 Local Convergence of Random Network Models

4.1 Local Convergence

Sometimes the neighbourhood around a random vertex converges to a common random graph structure as the graph grows large. This motivates the concept of local convergence, which is quite useful in simplifying the analysis of random graph properties. To begin with, we define the r-neighbourhoods of graphs.

Definition 4.1 (r-neighbourhood)**.** *Let G be a graph and $v \in V(G)$. Let r be a positive integer. The* ***r-neighbourhood*** *of v is*

$$B_r^{(G)}(v) = \{u \mid d_G(u, v) \leq r\} .$$

Now we present the definition of local convergence.

Definition 4.2 (Local Convergence)**.** *([2, p60]) Let (G_n, o_n) be a sequence of random rooted graphs where G_n is a random graph and o_n a uniformly randomly chosen vertex in G_n.*

- *We say* (G_n, o_n) ***converges locally weakly*** *to* (G, o) *if for every integer* $r > 0$ *and finite rooted graph* H^*,
$$\mathbb{P}\Big(B_r^{(G_n)}(o_n) \cong H^*\Big) \to \mathbb{P}\Big(B_r^{(G)}(o) \cong H^*\Big).$$
We denote this $(G_n, o_n) \xrightarrow{d} (G, o)$.
- *We say* (G_n, o_n) ***converges locally in probability*** *to* (G, o) *if for every integer* $r > 0$ *and finite rooted graph* H^*,
$$\mathbb{E}\left[\mathbb{1}_{\left\{B_r^{(G_n)}(o_n) \cong H^*\right\}} \middle| G_n\right] \xrightarrow{\mathbb{P}} \mathbb{P}\Big(B_r^{(G)}(o) \cong H^*\Big).$$
We denote this $G_n \xrightarrow{\mathbb{P}} (G, o)$.

With local convergence, quite a few properties of graphs will converge to their corresponding properties in the local limit, which greatly simplifies the analysis. To begin with, sizes of the r-neighbourhoods converges to the r-degree of vertex o in the limiting graph, which is clear from the definition of local convergence. Also, the sequence of local clustering coefficient converges in probability to the expectation of the coefficient of root o, making it a loteasier to compute.

Theorem 4.3 (Convergence of Clustering Coefficient)**.** *([2, p71]) Let* $(G_n)_{n>0}$ *be a sequence of graphs with* $|V(G_n)| \to \infty$. *Suppose* $G_n \xrightarrow{\mathbb{P}} (G, o)$. *Then*
$$\overline{CC}_{G_n} \xrightarrow{\mathbb{P}} \mathbb{E}\left(\frac{\Delta_G(o)}{(d_G(o)(d_G(o)-1))}\right)$$

4.2 Local Convergence of Random Network Models

Now we present the local convergence results related to the random network models introduced in section 3

Theorem 4.4 (Local Convergence of the Erdős–Rényi Model)**.** *([2, p63])* $\mathrm{ER}\big(n, \frac{\lambda}{n}\big)$ *converges locally in probability to a Poisson Branching process with offspring distribution* $\mathrm{Pn}(\lambda)$.

Theorem 4.5 (Locally Tree-like Nature of the Configuration Model)**.** *([2, p142]) Let* $G_n = \mathrm{CM}_n(\mathbf{d})$ *be a sequence of configuration graphs and* D_n *the degree of a randomly selected node in* $V(G_n)$. *Then under regularity conditions below,*

- $D_{G_n} \xrightarrow{d} D$ *for some random variable* D.
- $\mathbb{E}(D_{G_n}) \to \mathbb{E}(D) < \infty$.

G_n *converges locally in probability to the unimodular Galton-Watson tree* (G, o) *with root offspring distribution* $(\mathbb{P}(D = k))_{k \geq 0}$.

Theorem 4.6 (Local Convergence of Preferential Attachment Models)**.** *([2, p204]) Let* $m \geq 1$ *and* $\delta > -m$. *The preferential attachment model* $\mathrm{PA}_n^{(m,\delta)}(d)$ *converges locally in probability to the Pólya point tree.*

We will not go into the details of what the local limits are precisely, but note that all of the local limits are tree-like as illustrated by figure 4.1. This means the clustering coefficient of those random network models must converge to 0 in probability by theorem 4.3 because no triangles exist in a tree. In the next section (section 5) we will introduce and analyse a potential solution to this issue.

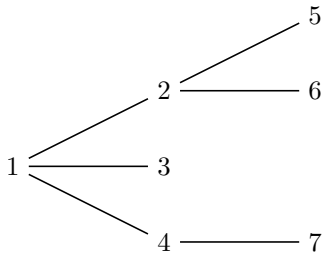


Figure 4.1: An example of a tree.

5 A General Embellishment

As we demonstrated in section 4, many common random graph models have a tree-like local limit. As a result, they cannot be highly clustered since the number of triangles is asymptotically small compared to the number of wedges.

We observe that the issue can be resolved by completing a sufficient number of triangles in the original graph. In this section, we will propose a generally applicable embellishment to improve the clustering coefficient of random graphs in subsection 5.1. Then we will show study some of its properties in subsection 5.2.

5.1 Construction

For any sequence of random graphs $G_n = (V(G_n), E(G_n))$, define

$$W(G_n) = \{\{u, v\} \mid d_{G_n}(u, v) = 2\}.$$

Define the embellished graph G'_n such that starting from G_n, edges in $W(G_n)$ are added independently with probability q (figure 5.1). We will consistently use G'_n to denote the random graph obtained by embellishing G_n with this operation.

In the context of a friendship network, this operation simulates the process of establishing friendship with an individual's friends-of-friends. Such connections will complete a considerable amount of triangles in the original graph, boosting the clustering coefficient as a result.

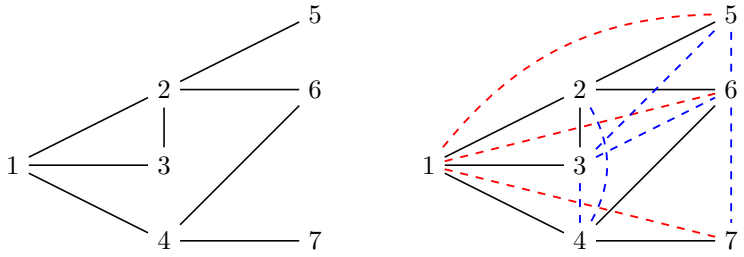


Figure 5.1: The neighbourhood of a graph (left) and the edges that are possibly added (right). The red edges corresponds to the added connections of node 1.

5.2 General Results

We introduce the notations we will use for the rest of this subsection. Let G be an arbitrary random graph. Define $d_G^{(2)}(v) = |\{u | d_G(u,v) = 2\}|$ the number of nodes of exactly distance 2 away from v. Define $D_G = d_G(V)$ to be the degree of a uniformly random vertex V in $V(G)$ and $D_G^{(2)} = d_G^{(2)}(V)$. Following the convention above, we use $d_{G'}(v)$ and $D_{G'}$ to denote the degrees in the embellished graph.

A critical observation is that knowing the structure of the 2-neighbourhood in the original graph is sufficient to construct the structure of the 1-neighbourhood of the embellished graph. This yields the relationship between D'_G and $\left(D_G, D_G^{(2)}\right)$:

Lemma 5.1 (Embellished Degree Distribution). *Let G be a random graph. Then conditioned on $\left(D_G, D_G^{(2)}\right)$,*

$$D'_G \stackrel{d}{=} D_G + X$$

where $X \sim \mathrm{Bi}\left(D_G^{(2)}, q\right)$ and X and D_G are independent conditioned on $\left(D_G, D_G^{(2)}\right)$.

Figure 5.1 illustrates the idea. The connections of a node in the embellished graph consists of connections in the original graph (black) and added connections (red). Knowing the original first-degree and second-degree distributions in the original graph allows us to specify the degree distribution of the embellished model.

Now suppose we have a sequence of random graphs $(G_n)_{n>0}$. We extend the above notation and use D_{G_n} to denote the typical degree of G_n, etc.

Theorem 5.2 (Weak Convergence of Embellished Degree). *Assume $\left(D_{G_n}, D_{G_n}^{(2)}\right) \xrightarrow{d} \left(D, D^{(2)}\right)$ for some random variable $\left(D, D^{(2)}\right)$. Then $D_{G'_n} \xrightarrow{d} D'$ where conditioned on $\left(D, D^{(2)}\right)$*

$$D' \stackrel{d}{=} D + X$$

where $X \sim \mathrm{Bi}\left(D^{(2)}, q\right)$ and X and D are independent conditioned on $\left(D, D^{(2)}\right)$.

This theorem basically follows from the previous lemma and we will leave the formal proof to subsection A.1.

Also observe that our operation has no effect on the component structure:

Theorem 5.3 (Component Structure Unchanged). *G' has the same component structure as G. In other words, u and v are connected in G' if and only if u and v are connected in G.*

Proof. This is obvious since our operation only adds edges within connected components and hence distinct components are not joint in the process. □

A direct consequence of this is that the highly-connected property is preserved under the embellishment:

Corollary 5.4 (Preservation of Highly-connected Graphs). *If $(G_n)_{n>0}$ is a sequence of highly-connected graphs then so is $(G'_n)_{n>0}$.*

Finally, we present a few results on typical distances in embellished graphs.

Lemma 5.5 (Bound on Graph Distance). *Let G be a graph (not random). Then for any pair of nodes $u, v \in V(G)$,*

$$\frac{1}{2} d_G(u,v) \leq d_{G'}(u,v) \leq d_G(u,v)\,.$$

Proof. The upper bound is clear since no edges are removed during the embellishment process.

For the lower bound we proceed by contradiction. Suppose there is a path (with positive probability) $\mathbf{w} = (w'_0, \ldots, w'_k)$ between u and v in the embellished graph G' with $k < \frac{1}{2}d_G(u,v)$. We will show that there must be path in G between u, v with length smaller than $d_G(u,v)$, which contradicts the premise. Notice that for any pair of nodes i, j, $\{i,j\} \in E(G')$ with positive probability if and only if $\{i,j\} \in E(G)$ or $\{i,j\} \in W(G)$. Therefore, $\{w_i, w_{i+1}\} \in E(G')$ with positive probability means $\{w_i, w_{i+1} \in E(G)\}$ or $\{w_i, w_{i+1} \in E(G)\} \in W(G)$. If $\{w_i, w_{i+1} \in E(G)\} \in E(G)$ then the distance between them is 1; otherwise $\{w_i, w_{i+1} \in E(G)\} \in W(G)$ and the $d_G(w_i, w_{i+1})$ between them is 2. This means $d_G(u,v)$ must be no greater than the length of the longest path we can derive from $\mathbf{w}$, which is of length

$$d(w_0, w_2) + d(w_1, w_2) + \cdots + d(w_{k-1}, w_k) = 2d_{G'}(u,v) < d_G(u,v)\,,$$

and the proof is complete. □

Theorem 5.6 (Stochastic Domination of Typical Distance)**.** *L_G stochastically dominates $L_{G'}$. That is, for any l,*

$$\mathbb{P}(L_G \geq l) \geq \mathbb{P}(L_{G'} \geq l)\,.$$

Proof. For a random graph G, we generate the coupling $(\bar{L}_G, \bar{L}_{G'})$ with G' the embellished graph of G. Since $\mathbb{P}(\bar{L}_{G'} \leq \bar{L}_G) = 1$ by lemma 5.5, the result follows. □

Corollary 5.7 (Preservation of Small-world Properties)**.** *If $(G_n)_{n>0}$ is a sequence of small-world graphs then so is $(G'_n)_{n>0}$. If $(G_n)_{n>0}$ is a sequence of ultra-small world graphs then so is $(G'_n)_{n>0}$.*

Proof. Using theorem 5.6 and the assumption, there exists a $K > 0$ such that

$$\mathbb{P}\big(L_{G'_n} \leq K \log n\big) \leq \mathbb{P}\big(L_{G'_n} \leq K \log n\big) \leq 1\,.$$

Taking limits as $n \to \infty$ shows $\mathbb{P}\big(L_{G'_n} \leq K \log n\big) \to 1$. The ultra-small world version can be shown in a similar way. □

5.3 Local Convergence of Embellished Models

Here we give stronger results that can be shown when the original sequence of random graphs converges locally. Our first result shows that local convergence of random graphs is preserved under the embellishment.

Theorem 5.8 (Local Convergence is Preserved)**.** *Let G_n be a sequence of random graphs. Then if $(G_n, o_n) \xrightarrow{d} (G, o)$ then $(G'_n, o'_n) \xrightarrow{d} (G', o)$.*

We will leave the proof the theorem in subsection A.2.

Theorem 5.9 (Degrees of Embellished Graphs)**.** *Let G_n be a sequence of random graphs with $(G_n, o_n) \xrightarrow{d} (G, o)$. Then $d_{G'_n}(o_n) \xrightarrow{d} d_{G'}(o)$ where $d_{G'}(o) \mid d_G(o), d_G^{(2)}(o) \overset{d}{=} d_G(o) + X$ where $X \mid d_G(o), d_G^{(2)}(o) \sim Bi\Big(d_G^{(2)}(o), q\Big)$.*

Proof. As illustrated by figure 5.1, the only source of degree change is from the added edges (red), which follows $Bi\Big(d_G^{(2)}(o), q\Big)$ since edges are iid Bernoulli variables with parameters q. □

The following theorem shows that our embellishment boosts the clustering coefficient.

Theorem 5.10 (Clustering Coefficient Boosted)**.** *Let* $(G_n)_{n>0}$ *be a sequence of graphs with* $|V(G_n)| \to \infty$*. Suppose* $G_n \xrightarrow{\mathbb{P}} (G, o)$ *and* $\mathbb{P}(d_G(o) > 1) > 0$*. Then*

$$\overline{\mathrm{CC}}'_{G_n} \to \mathbb{E}\left(\frac{\Delta_G(o)}{d_G(o)(d_G(o)-1)}\right) > 0\,.$$

Here $\mathbb{P}(d_G(o) > 1) > 0$ is a non-degeneracy condition to ensure the original graph does not consist of isolated edges or vertices only, in which case no wedges present in the graph and the clustering coefficient is undefined. This theorem justifies the boost in the clustering coefficient. We will leave the proof to subsection A.3.

6 Discussion

In this paper we introduced random networks and some of the properties of real-world networks. Despite being consistent with most features of real-world networks, many existing random networks fail to explain the highly-clustered property, which motivates this study. We introduced an embellishment that rectifies this problem. We showed that it preserves the local convergence of random networks and boosts their clustering coefficient to a positive constant, no matter what it was previously.

In general our embellishment performs well. It is generally applicable and it boosts the clustering coefficient of many common random network models.

There are also a few drawbacks with our embellishment. It tends to increase the degree of vertices with a high number of second-order neighbours significantly, which may disrupt the degree distribution in some extreme cases. This makes the effect of the embellishment heavily dependant on the second-order neighbourhood of the original graph, which might be complicated.

Recent research shows that scale-free graphs with exponent $1 < \gamma < 2$ cannot be highly-clustered in the sense that the clustering coefficient must converge to 0 [3]. This means our embellishment will likely break the scale-free property of random graphs. However, whether the scale-free property is preserved for larger exponents remain unclear.

Further research may be done by running simulations to confirm our theorems and perhaps reveal other properties of this embellishment.

7 Acknowledgement

I would like to express my deepest appreciation to my supervisor Dr Nathan Ross, for his patient guidance and inspiring suggestions throughout the vacation scholarship program.

8 References

References

[1] Remco van der Hofstad. *Random Graphs and Complex Networks*. Vol. 1. 2018. URL: https://www.win.tue.nl/~rhofstad/NotesRGCN.html.

[2] Remco van der Hofstad. *Random Graphs and Complex Networks*. Vol. 2. 2021. URL: https://www.win.tue.nl/~rhofstad/NotesRGCN.html.

[3] Liudmila Ostroumova Prokhorenkova. *Global clustering coefficient in scale-free weighted and unweighted networks*. 2015. arXiv: 1507.00925 [cs.SI].

A Appendix

A.1 Proof of Theorem 5.2

Theorem 5.2 (Weak Convergence of Embellished Degree). *Assume* $\left(D_{G_n}, D^{(2)}_{G_n}\right) \xrightarrow{d} \left(D, D^{(2)}\right)$ *for some random variable* $\left(D, D^{(2)}\right)$. *Then* $D_{G'_n} \xrightarrow{d} D'$ *where conditioned on* $\left(D, D^{(2)}\right)$

$$D' \stackrel{d}{=} D + X$$

where $X \sim \mathrm{Bi}\left(D^{(2)}, q\right)$ *and* X *and* D *are independent conditioned on* $\left(D, D^{(2)}\right)$.

Proof. Expand using law of total probability,

$$\mathbb{P}\left(D_{G'_n} = k\right) = \sum_{k_1,k_2=0}^{\infty} \mathbb{P}\left(D_{G'_n} = k \middle| D_{G_n} = k_1, D^{(2)}_{G_n} = k_2\right)\mathbb{P}\left(D_{G_n} = k_1, D^{(2)}_{G_n} = k_2\right). \tag{A.1}$$

Here we apply lemma 5.1 to the conditional probability,

$$\begin{aligned}
\mathbb{P}\left(D_{G'_n} = k \middle| D_{G_n} = k_1, D^{(2)}_{G_n} = k_2\right) &= \mathbb{P}\left(D_{G_n} + \mathrm{Bi}(D^{(2)}_{G_n}, q) = k \middle| D_{G_n} = k_1, D^{(2)}_{G_n} = k_2\right) \\
&= \mathbb{P}\left(k_1 + \mathrm{Bi}(k_2, q) = k \middle| D_{G_n} = k_1, D^{(2)}_{G_n} = k_2\right) \\
&= \mathbb{P}\left(k_1 + \mathrm{Bi}(k_2, q) = k \middle| D = k_1, D^{(2)} = k_2\right) \\
&= \mathbb{P}\left(D + \mathrm{Bi}(D^{(2)}, q) = k \middle| D = k_1, D^{(2)} = k_2\right) \\
&= \mathbb{P}\left(D' = k \middle| D = k_1, D^{(2)} = k_2\right).
\end{aligned} \tag{A.2}$$

For simplicity, we abused the notation to work with Bi directly. The other term in A.1 converges due to the assumption $(D_n, D_n * (2)) \xrightarrow{d} \left(D, D^{(2)}\right)$,

$$\mathbb{P}\left(D_{G_n} = k_1, D^{(2)}_{G_n} = k_2\right) \to \mathbb{P}\left(D = k_1, D^{(2)} = k_2\right) \tag{A.3}$$

as $n \to \infty$ Using A.2 and A.3, the result follows:

$$\mathbb{P}\left(D_{G'_n} = k\right) \to \sum_{k_1,k_2=0}^{\infty} \mathbb{P}\left(D' = k \middle| D = k_1, D^{(2)} = k_2\right)\mathbb{P}\left(D = k_1, D^{(2)} = k_2\right) = \mathbb{P}(D' = k).$$

□

A.2 Proof of Theorem 5.8

Theorem 5.8 (Local Convergence is Preserved). *Let* G_n *be a sequence of random graphs. Then if* $(G_n, o_n) \xrightarrow{d} (G, o)$ *then* $(G'_n, o'_n) \xrightarrow{d} (G', o)$.

Proof. We need to show that for any positive integer r,

$$\mathbb{P}(B_r^{(G'_n)}(o'_n) \cong H^*) \to \mathbb{P}(B_r^{(G')}(o) \cong H^*).$$

Since $\mathbb{P}(B_r^{(G_n)}(o_n) \cong H^*) \to \mathbb{P}(B_r^{(G)}(o) \cong H^*)$ for any r, we can make a coupling $(\bar{B}_r^{(G_n)}(o_n), \bar{B}_r^{(G)}(o))$ such that $\mathbb{P}(\bar{B}_r^{(G_n)}(o_n) \ncong \bar{B}_r^{(G)}(o)) \to 0$. With this coupling,

$$\mathbb{P}\left(B_r^{(G'_n)}(o'_n) \cong H^*\right) = \mathbb{P}\left(B_r^{(G'_n)}(o'_n) \cong H^*, \bar{B}_{2r}^{(G_n)}(o_n) \cong \bar{B}_{2r}^{(G)}(o)\right) + \mathbb{P}\left(B_r^{(G'_n)}(o'_n) \cong H^*, \bar{B}_{2r}^{(G_n)}(o_n) \ncong \bar{B}_{2r}^{(G)}(o)\right). \tag{A.4}$$

Notice that the second term in (A.4) vanishes as $n \to \infty$:

$$\mathbb{P}\left(B_r^{(G'_n)}(o'_n) \cong H^*, \bar{B}_{2r}^{(G_n)}(o_n) \ncong \bar{B}_{2r}^{(G)}(o)\right) \leq \mathbb{P}\left(\bar{B}_{2r}^{(G_n)}(o_n) \ncong \bar{B}_{2r}^{(G)}(o)\right) \to 0\,. \tag{A.5}$$

For the first term we are under the event $\left\{\bar{B}_{2r}^{(G_n)}(o_n) \cong \bar{B}_{2r}^{(G)}(o)\right\}$. Since their structures are isomorphic, we may couple the Bernoulli variables generating their embellished versions and obtain a coupling $\left(\bar{B}_r^{(G'_n)}(o_n), \bar{B}_r^{(G')}(o)\right)$ such that $\bar{B}_r^{(G'_n)}(o_n) \cong \bar{B}_r^{(G')}(o)$. As $n \to \infty$ this gives

$$\begin{aligned}
&\mathbb{P}\left(B_r^{(G'_n)}(o'_n) \cong H^*, \bar{B}_{2r}^{(G_n)}(o_n) \cong \bar{B}_{2r}^{(G)}(o)\right) \\
=&\mathbb{P}\left(\bar{B}_r^{(G'_n)}(o'_n) \cong H^*, \bar{B}_{2r}^{(G_n)}(o_n) \cong \bar{B}_{2r}^{(G)}(o)\right) \\
=&\mathbb{P}\left(\bar{B}_r^{(G')}(o) \cong H^*, \bar{B}_{2r}^{(G_n)}(o_n) \cong \bar{B}_{2r}^{(G)}(o)\right) \\
=&\mathbb{P}\left(\bar{B}_r^{(G')}(o) \cong H^*\right) - \mathbb{P}\left(\bar{B}_r^{(G')}(o) \cong H^*, \bar{B}_{2r}^{(G_n)}(o_n) \ncong \bar{B}_{2r}^{(G)}(o)\right) \\
\to&\mathbb{P}\left(\bar{B}_r^{(G')}(o) \cong H^*\right) \\
=&\mathbb{P}\left(B_r^{(G')}(o) \cong H^*\right)
\end{aligned} \tag{A.6}$$

where the convergence is obtained similarly to (A.5). Combining the result of (A.5) and (A.6), the desired result follows from (A.4). □

A.3 Proof of theorem 5.10

Theorem 5.10 (Clustering Coefficient Boosted)**.** *Let $(G_n)_{n>0}$ be a sequence of graphs with $|V(G_n)| \to \infty$. Suppose $G_n \xrightarrow{\mathbb{P}} (G, o)$ and $\mathbb{P}(d_G(o) > 1) > 0$. Then*

$$\overline{\mathrm{CC}}'_{G_n} \to \mathbb{E}\left(\frac{\Delta_G(o)}{d_G(o)(d_G(o)-1)}\right) > 0\,.$$

Proof. By Theorem 4.3,

$$\overline{\mathrm{CC}}_{G'_n} \to \mathbb{E}\left(\frac{\Delta_{G'}(o)}{d_{G'}(o)(d_{G'}(o)-1)}\right).$$

To show the quotient is positive, note that both the numerator and denominator are always non-negative. Hence it suffice to find a realisation of the quotient that is positive with the realisation occurring with a positive probability. This is clear since we assume $\mathbb{P}(d_G(o) > 1) > 0$, so there exists a realisation $d_G(o) = k > 1$ with positive probability. For this k, one can check that adding one connection between the neighbours of o is a possible realisation that guarantees a positive quotient since $\Delta_{G'}(o) \geq 2$ in this case. □