ITERATED MODELS FOR SOCIAL NETWORKS

by

Erin Kathleen McKenna Meger

M.Sc. Applied Mathematics, Ryerson University, 2016

B.Sc. Mathematics, Wilfrid Laurier University, 2014

A dissertation presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the program of

Mathematical Modelling and Methods

Toronto, Ontario, Canada, 2020

© Erin Meger, 2020

Author's Declaration for Electronic Submission of a Dissertation

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

Abstract

Erin Kathleen McKenna Meger

Iterated Models for Social Networks

PhD in Mathematical Modelling and Methods, 2020

Ryerson University

We define two novel iterative models of social networks. The models are deterministic processes that generate graphs over discrete time-steps, and the properties of these graphs will be explored. The first model generalizes two known models: the Iterated Local Transitivity Model and the Iterated Local Anti-Transitivity Model. The Iterated Local Model includes as input an infinite binary sequence that determines the way the graphs are constructed over time. These models each utilize the underlying graph structure in previous time-steps. Subsequently, we define a model that is independent of the structure of the graph at the previous time-step. The Iterated Global Model creates new adjacencies based on subsets of vertices of a prescribed cardinality.

We prove complex network properties of the Iterated Local Model, such as the small-world property and bad spectral expansion. We also present graph-theoretic properties of the model, such as bounds on the chromatic number, domination number, and Hamiltonicity properties. Analogously, for the Iterated Global Model, we prove both complex network and graph-theoretic properties.

Acknowledgements

Contents

List of Figures

CHAPTER 1

Complex networks and graph theory

1.1. Introduction

Over the last two decades, research in modelling complex networks has become of great interest to mathematicians and theoretical computer scientists. The emergence of the study of complex networks such as the web graph and on-line social networks has focused attention on these large-scale graphs and their properties [**4, 7, 11, 19, 20, 21, 24**]. The theory of complex networks analyzes graph-theoretic properties in real-world networks arising in technological, social, and biological contexts.

In this thesis, we will begin by including an introduction to graph theory. This will include all requisite background for future chapters. The study of complex networks was preceded by the study of random graph models, so we will explore some of the influential literature in the area. We include a discussion of the properties of complex networks that will be examined throughout the rest of the thesis and present well-known models for complex networks. We conclude with a summary of the thesis.

1.2. Graph theory

We define a *graph* $G = (V, E)$ as a pair of a *vertex set* $V = V(G)$, an *edge set* $E = E(G)$ containing pairs of vertices. The *order* of a graph G is $|V(G)|$, and its *size* is $|E(G)|$; when

these are finite, then we say G is finite. We write uv if vertices u and v form an edge, and say that u and v are *adjacent*, and we call them *neighbours*. We say that for a vertex, v, the neighbour set of v, denoted $N(v)$, is the set of all vertices adjacent to v. Additionally, we use the notation $N[v]$ to denote the set of $N(v) \cup \{v\}$, called the *closed neighbour set*. We say u and v are *incident* with the edge uv, and that u and v are *endpoints* of the edge uv. The *degree* of v, denoted by $\deg_G(v)$, is the number of edges incident with v. The subscript G may be omitted when there is no risk of confusion. Notice that $\deg_G(v) = |N(v)|$. The *minimum degree* of G is $\delta(G) = \min\{\deg(v) : v \in V\}$, and the *maximum degree* of G is $\Delta(G) = \max\{\deg(v) : v \in V\}$. A vertex that has degree zero is called an *isolated vertex*, while a vertex adjacent to all others is called a *universal vertex*.

Graphs are often depicted by their drawings; see Figure 1.1.
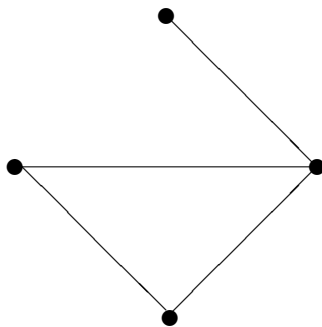
Figure 1.1. An example of a graph of order and size 4.

Unless otherwise stated, we only consider finite, *simple graphs* which exclude edges of the form xx, called loops, or multiple edges between the same pair of two vertices. A graph

is *directed* if each of its edges has an orientation, and we can consider the edge-set as a set of ordered pairs of vertices rather than unordered pairs. Unless otherwise stated, we will not consider directed graphs in this thesis.

We now recall the *First Theorem of Graph Theory*, which provides an elementary but important relationship between the degrees and the size of a graph. The theorem and its proof are folklore.

THEOREM 1.1. *If $G = (V, E)$ is a graph, then*

$$\sum_{u \in V} \deg(u) = 2|E| \ .$$

PROOF. Each edge is counted twice, once for each endpoint. $\square$

The previous theorem indicates that all graphs have an even degree sum, since the number of edges is always an integer. This gives the following corollary.

COROLLARY 1.2. *In a graph, the number of vertices of odd degree is even.*

A *subgraph* H of a graph G is a graph such that $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. We say a subgraph H is an *induced subgraph* of G when $uv \in E(H)$ if and only if u and $v \in V(H)$. We use the notation $G[S]$, where $S \subseteq V(G)$ to represent the subgraph of G induced by the set of vertices S. We demonstrate this concept in Figure 1.2. On the left, the graph G has vertices $V(G) = \{1, 2, 3, 4\}$ and on the right, we see the subgraph induced by $\{2, 3, 4\}$. A subgraph H is *spanning* in G if $V(H) = V(G)$.

3
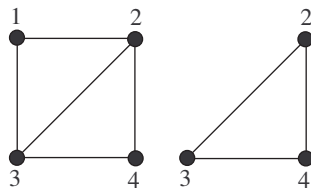
FIGURE 1.2. An example of an induced subgraph.

A graph G is a *clique* if G contains all possible edges between its vertices. The *complement* of a graph $G = (V, E)$ is $\overline{G} = (V, \{xy : xy \notin E\})$. An *independent set* is the complement of a clique; that is, a set of vertices with no edges between them. A graph $G = (V, E)$ is called *k-partite* if V can be partitioned into k independent sets. If $k = 2$, then we say G is *bipartite*. Two sets $A, B \subseteq V(G)$ are said to *partition* the vertex set whenever $A \cup B = V(G)$ and $A \cap B = \emptyset$.

A *path* on $k + 1$ vertices is an ordered set of vertices $\{v_0, v_1, \ldots v_k\}$ such that each v_i is adjacent to v_{i+1} and $v_i \neq v_j$ for all $i \neq j$; a *cycle* on k vertices is a path on $k + 1$ vertices with $v_0 = v_k$ and $v_i \neq v_j$ for all $i \neq j$. The path of order n is denoted by P_n. We use the notation C_n for a cycle of order n. We say a graph G is *connected* whenever there exists a path between u and v for all $u, v \in V(G)$. The *components* of a graph G are the maximal connected induced subgraphs of G, with respect to set inclusion.

The length of a shortest path in G between u and v is called the *distance* between u and v, denoted by $\text{dist}_G(u, v)$. When the graph in which we are finding the distance is clear from context, we will omit the subscript. The *diameter* of G is determined as follows.

4

$$\text{diam}(G) = \max\{\text{dist}(v, w) : v, w \in V(G)\}$$

The *radius* of a graph is the largest integer r such that every vertex has at least one vertex at a distance r. For all graphs G the radius is at most the size of the diameter.

The depicted graph G, in Figure 1.3, $\text{diam}(G) = 4$, $\text{dist}_G(a, h) = 4$,, $\text{dist}_G(c, d) = 2$, and the radius is 2. A shortest path between two vertices may not exist whenever the graph is not connected, in which case we use ∞ to denote their distance.
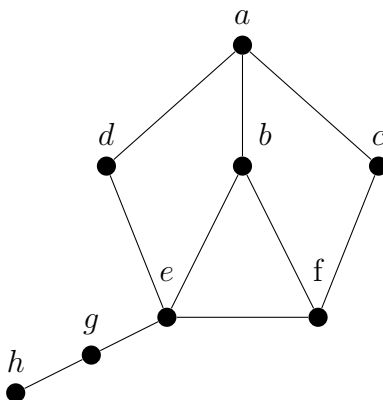
FIGURE 1.3. A simple graph used to illustrate diameter and distance.

For a positive integer k, a graph G is *k-connected* if there does not exist a set of $k - 1$ vertices whose removal disconnects the graph. A *cut-vertex* is one whose removal disconnects the graph. Note that a graph is 1-connected precisely when it contains a cut-vertex.

We say that two graphs, say G and H, are *isomorphic*, written $G \cong H$, if there is a bijection $f : V(G) \longrightarrow V(H)$ such that $uv \in E(G)$ if and only if $f(u)f(v) \in E(H)$. See Figure 1.4 for two isomorphic graphs.
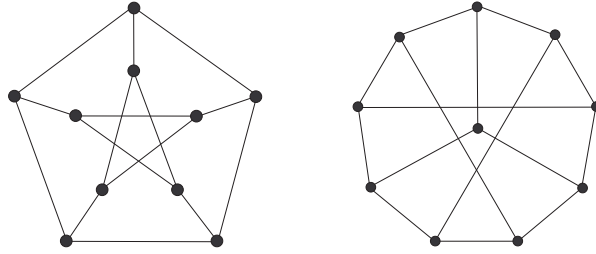
FIGURE 1.4. Two graphs isomorphic to the well-known Petersen graph.

A connected graph with no cycle is a *tree*. The next theorem characterizes trees and is included for completeness, the proof of which can be found in the literature [**23**].

THEOREM 1.3. *The following are equivalent.*

(1) *The graph G is a tree.*

(2) *The graph G is connected and $|E(G)| = |V(G)| - 1$.*

(3) *There is a unique path between every pair of vertices in G.*

1.2.1. Graph Parameters. We close this section with a brief sampling of important graph parameters that will be discussed throughout the thesis, but particularly in Chapter 4. Let $G = (V, E)$ be a graph. We call $S \subseteq V$ a *dominating set* for G if for all $v \notin S$, there exists $w \in S$ such that $vw \in E$. The minimum cardinality of all dominating sets in G is denoted $\gamma(G)$, and is called the *domination number* of G.

To colour a graph, we assign a colour from the set $\{c_0, c_1, \ldots, c_k\}$ to each vertex. A *proper colouring* is achieved when no two neighbouring vertices have the same colour. The *chromatic number* of a graph G, denoted by $\chi(G)$, is the minimum number of colours required to achieve a proper colouring of G. If G can be coloured using at most k colours, then we say

6

that G is *k-colourable.* The maximum order of a clique in G is called the *clique number* of G, denoted by $\omega(G)$. The maximum order of an independent set in G is called the *independence number* of G, denoted by $\alpha(G)$.

We consider the graph G in Figure 1.5. The set $\{a, e\}$ is a dominating set, and is, in fact, the smallest dominating set; hence, the domination number of this graph is two. The chromatic number is $\chi(H) = 3$ and $\omega(G) = 3$. By inspection, the maximum order of an independent set is two.
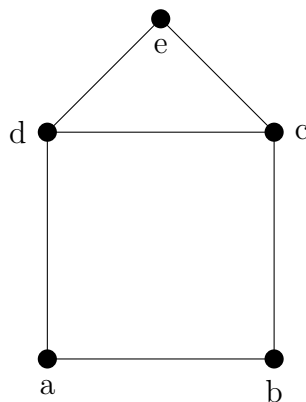
FIGURE 1.5. An example of graph parameters

A *Hamilton path* is a path P in G such that each vertex of G is visited along the path exactly once. A *Hamilton Cycle* is a Hamilton path that has its final vertex as its initial vertex. If a graph G has a Hamilton cycle, then we say the graph is *Hamiltonian.* Determining whether a graph has a Hamilton cycle is a deep problem within structural graph theory, and many problems remain open [**23**].

1.2.2. Asymptotic notation. We finish the section with asymptotic notation. Let $f(n)$ and $g(n)$ be two functions whose domain is some fixed subset of $\mathbb{R}$, and assume that

$g(n) > 0$ for all n. We say that f *is of order at most* g, written $f(n) = O(g(n))$, if there exist constants $A > 0$ and $N > 0$ such that for all $n > N$, we have that

$$|f(n)| \leq A|g(n)|.$$

We say that f *is of order at least* g, written $f(n) = \Omega(g(n))$, if there exist constants $A > 0$ and $N > 0$ such that for all $n > N$,

$$f(n) \geq Ag(n).$$

We say that f *is of order* g, written $f(n) = \Theta(g(n))$, if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

We say that f is *of order smaller than* g, written $f(n) = o(g(n))$ if

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 0.$$

The function f is *of order larger than* g, written $f(n) = \omega(g(n))$ if

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = \infty.$$

Finally, f is *asymptotically equal* to g, written $f(n) \sim g(n)$ or $f(n) = (1 + o(1))g(n)$, if

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 1.$$

1.3. Complex networks

The study of complex networks is an emerging field in discrete mathematics and theoretical computer science. Models of these networks may be used to predict the structure and evolution of protein-protein interaction networks, on-line social networks, and the web graph [14, 19]. These networks have four key properties: they are large scale, evolve over time, have power law degree distributions, and satisfy the small-world property [13]. An early stochastic network model was the Erdős-Rényi random graph [2, 7, 13, 12]. Erdős-Rényi random graphs fell short when it came to studying complex networks; with high probability, these graphs exhibit a binomial degree distribution rather than a power law degree distribution. Early complex networks were designed using the premise of preferential attachment [3, 6, 13]. Recently, two new deterministic models, the Iterated Local Transitivity (ILT) model and the Iterated Local Anti-Transitivity (ILAT) model, were introduced with the aim of simulating social networks [9, 10].

In thIS thesis, we will build upon the ILT and ILAT models and study complex networks using new deterministic methods. Each of the ILT and ILAT models separately examines the ideas of transitivity and anti-transitivity. These models are a mathematical reflection of the folkloric adages "friends of friends are more likely friends" and "the enemy of my enemy is my friend." The model we describe in Chapter 3 will incorporate both transitivity and anti-transitivity sequentially over discrete time-steps.

1.3.1. Complex Networks and Their Properties. There is no precise definition of a complex network as they are real-world graphs that are constantly evolving. However, there are four key properties that most researchers agree upon: large-scale, evolving over time, power-law degree distribution, and small-world property [**7, 13**].

In this more heuristic context, large-scale depends on the type of network we are considering. For instance, the web graph has trillions of vertices, whereas protein-protein interaction networks have thousands of vertices [**7**]. Both of these networks are considered to be complex networks. Networks that evolve over time are those which exhibit the creation, and possibly the deletion, of vertices over time. This is also seen in some network models such as the ACL Preferential Attachment model that will be discussed subsequently.

In complex networks, we observe that there are many vertices of low degree and only a few vertices of high degree. This property can also be found among social networks such as Facebook [**7**].

The parameter $N_{k,G}$ is *number of vertices of degree k in G*. For simplicity, suppose that $|V(G)| = t$. Hence, $N_{k,G}$ is an integer in the interval $[0, t]$. The *degree distribution* of G is the sequence $(N_{k,G} : 0 \leq k \leq t)$. We say that the degree distribution of G follows a *power law* if for each degree k, we have as follows for a fixed real constant $\beta > 1$.

$$\frac{N_{k,G}}{t} \sim k^{-\beta}, \tag{1}$$

Informally, G has many vertices of low degree, and a non-negligible number of high degree vertices. Such distributions are sometimes called *heavy-tailed distributions*, since the real-valued function $f(k) = k^{-\beta}$ exhibits a polynomial rather than exponential decay to 0 as k tends to infinity. We say that β is the *the power law exponent*. If G possesses a power law degree distribution, then we say G is a *power law graph*. Observe that if we take logarithms on both sides of (1), then the relationship is expressed as

$$\log(N_{k,G}) \sim \log(t) - \beta \log(k).$$

Hence, in the log-log plot, we obtain a straight line with slope $-\beta$. In the web graph, power law degree distributions were reported in Broder et al. [**11**]. Power laws have also been reported in social networks such as Twitter and in certain biological networks [**7**].

The fourth property for complex networks we discuss is the *small-world property*. Small-world networks have low distance and high local clustering. To make this more precise, we define the *average distance* of a graph G as follows:

$$L(G) = \frac{W(G)}{\binom{|V(G)|}{2}}, \qquad \text{where} \qquad W(G) = \sum_{u,v \in V(G)} \text{dist}_G(u,v)$$

The term $W(G)$ is called the *Wiener index*, which is the sum of all distances between all pairs of vertices. We will use the notation $G[N_G(x)]$ to represent the subgraph of G induced by the set of vertices in the neighbourhood of x. The *clustering coefficient* of G is defined by

$$C(G) = \frac{1}{|V(G)|} \sum_{x \in V(G)} c(x), \qquad \text{where} \qquad c(x) = \frac{\left|E\big(G[N_G(x)]\big)\right|}{\binom{\deg(x)}{2}}$$

Note that $L(G)$ and $C(G)$ are rational numbers in $[0,1]$. For a graph G of order n to satisfy the small-world property, we require that $L(G) = O(\log \log n)$ or $\mathrm{diam}(G) = O(\log n)$, and that $C(G)$ be larger than the clustering coefficient of a binomial random graph with the same average degree. We will discuss binomial random graphs further in Subsection 1.3.2.

There are two additional properties that are also explored when discussing complex networks, which have been demonstrated in many applications of complex networks, but were not originally included in the four main properties [**7, 13, 14, 16, 17**]. These are densification and bad spectral expansion.

We say a complex network *densifies* if the limit of the ratio of edges to vertices is unbounded [**17**]. Since complex networks evolve over time, we use G_t to denote the particular graph at time t [**7, 9, 10**].

$$\lim_{t \to \infty} \frac{|E(G_t)|}{|V(G_t)|} \to \infty$$

The *adjacency matrix* of a graph, G is a matrix representation of the edges. We consider each row and column in the matrix to represent each of the vertices $V(G) = \{v_1, ..v_n\}$ and there is a 1 in the entry a_{ij} of the matrix whenever vertex v_i and v_j are adjacent. The *spectral expansion* of a graph is the difference between the first two eigenvalues of the

12

adjacency matrix. In many real-world complex networks, the gap between these eigenvalues is quite large. This is known as *bad spectral expansion* [**16**].

One goal when studying complex networks is to simulate many or all of these properties, and thus we will consider them for each model defined in this thesis.

1.3.2. Graph Models. We will begin by discussing the Erdős-Rényi model for generating random graphs. Let $0 \leq p \leq 1$ and let Ω be the family of all graphs on n vertices. To every graph $G \in \Omega$ we assign a probability measure:

$$\mathbb{P}(\{G\}) = p^{|E(G)|}(1-p)^{\binom{n}{2}-|E(G)|}.$$

We denote this probability space by $\mathbb{G}(n,p)$. The space $\mathbb{G}(n,p)$ is often referred to as the *binomial random graph* or *Erdős-Rényi random graph*. Note also that this probability space can informally be viewed as a result of $\binom{n}{2}$ independent coin flips, one for each pair of vertices u, v, where the probability of success is equal to p. We also observe that if $p = 1/2$, then $\mathbb{P}(\{G\}) = 2^{-\binom{n}{2}}$ for any graph G on n vertices. Observe that $p = p(n)$ may be a function of n, but we restrict p to a constant value in this brief survey of the topic.

Binomial random graphs are not appropriate models for simulating complex networks, despite the fact they exhibit quite short distances [**13**]. In fact, we will often compare models to binomial random graphs especially in the case of clustering and densification. We note that the proofs of the following theorems are a part of folklore and may be found in the

literature [**2, 7**]. An event A_n in a probability space holds *asymptotically almost surely* or *a.a.s.* if $\mathbb{P}(A_n) \to 1$ as $n \to \infty$.

THEOREM 1.4. *For fixed $p \in (0,1)$, a.a.s. G in $\mathbb{G}(n,p)$ has diameter 2.*

Despite the conclusion of Theorem 1.4, the degree distribution of $\mathbb{G}(n,p)$ is binomial and not power law as we require for complex networks.

THEOREM 1.5. *A.a.s. for G in $\mathbb{G}(n,p)$, the degree of a vertex v in G equals*

$$pn + O(\sqrt{pn} \log n) = (1 + o(1))pn$$

1.3.3. Preferential Attachment Models. The first model of complex networks we will examine is the *preferential attachment* or *PA* model, discovered by Barabási and Albert in 1999 and made rigorous in 2001 by Bollobás, Riordan, Spencer, and Tusnády [**6, 3**].

As input, we fix $m \in \mathbb{N}$. The model generates graphs randomly over discrete time-steps, indexed by non-negative integers $t \geq 0$. At time-step $t = 0$, we add a single edge.

In the time-step $t + 1$, we add m edges from a new vertex v_{t+1} to existing vertices v_s to form the graph G_{t+1}. The edge $v_{t+1}v_s$ is added with probability

$$\frac{\deg_{G_t}(v_s)}{2(mt+1)}.$$

Note that we tacitly use here Theorem 1.1, as the number of edges at time-step t is $mt + 1$. We refer to this as the PA_m model. The term "preferential attachment" has to do with

the fact that vertices of higher degree are more likely to gain new neighbours in further time-steps; thus, the higher the degree of v_s, the more likely $v_{t+1}v_s$ is an edge. In this model we do not restrict to simple graphs and may exhibit multiple edges. The PA model generates graphs with power law degree distributions and the small-world property, as the next theorems demonstrate.

THEOREM 1.6 ([**6**]). *For all $m \geq 0$, and for any non-negative integer k satisfying $0 \leq k \leq t^{1/15}$ for time-step t in the PA_m model, we have that a.a.s.*

$$\frac{N_{k,t}}{t} = \frac{1 + o(1)}{k^3},$$

where $N_{k,t}$ is the number of vertices of degree k at time-step t.

THEOREM 1.7 ([**5**]). *For all $m \geq 0$, the diameter of the PA_m model graph at time-step t is a.a.s.*

$$(1 + o(1))\frac{\log t}{\log \log t} \ .$$

We reference one additional model, referred to as the *Aiello-Chung-Lu* or *ACL PA* model. It is named after William Aiello, Fan Chung, and Linyuan Lu, who discovered the model in 2001 [**1, 13**].

As input, we fix $p \in (0, 1)$. At time-step $t = 0$, the initial graph G_0 consists of a single vertex with a loop. A *vertex-step* is defined by adding a new vertex v and an edge uv where u is chosen from existing vertices by preferential attachment. An *edge-step* adds an edge uv,

15

where u and v are existing vertices both chosen by preferential attachment. To form the graph G_{t+1} from G_t, take a vertex-step with probability p and an edge-step with probability $1 - p$. We refer to this model as the $\mathcal{G}_{\mathrm{ACL}}(p)$ model.

By definition, the number of vertices in the ACL PA model is a random variable. Using the Chernoff bounds, it can be shown that $|V(G_t)|$ is concentrated around its expected value $1 + pt$ [**2**].

The ACL model a.a.s. generates power law graphs, as the next theorem demonstrates. We denote the Gamma function by Γ [**12**], defined originally by Bernoulli to generalize the factorial to complex numbers. We will use the following definition [**12**].

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$$

THEOREM 1.8 ([**13**]). *In* $\mathcal{G}_{\mathrm{ACL}}(p)$,

$$\lim_{t \to \infty} \mathbb{E}\left(\frac{N_{1,t}}{t}\right) = \frac{2p}{4 - p},$$

and for all $k > 1$,

$$
\begin{aligned}
\lim_{t \to \infty} \mathbb{E}\left(\frac{N_{k,t}}{t}\right) &= \frac{2p}{4 - p} \frac{\Gamma(k)\Gamma\left(2 + \frac{2}{2-p}\right)}{\Gamma\left(k + 1 + \frac{2}{2-p}\right)} \\
&= O\left(k^{-\left(2 + \frac{p}{2-p}\right)}\right).
\end{aligned}
$$

A concentration result for the ACL model can be derived using generalized martingale inequalities [**13**]. This result, together with Theorem 1.8 proves rigorously that $\mathcal{G}_{\text{ACL}}(p)$ generates graphs with power law degree distributions, with power law exponent $\beta = 2 + \frac{p}{2-p}$. Observe that by choosing p appropriately, the power law exponent can be any real number $2 \leq \beta \leq 3$.

1.4. Summary of the thesis

The first chapter motivated the discussion on complex network models and gave background on graph theory and properties of complex networks. In Chapter 2, we will provide background on the Iterated Local Transitivity model and the Iterated Local Anti-Transitivity model.

The new Iterated Local Model (ILM) will be introduced in Chapter 3, where we will prove results pertaining to its complex network properties. We will show that ILM graphs satisfy the small-world property and densification. The spectral gap will be discussed and it will be shown that the model has bad spectral expansion as found in real-world social networks. Chapter 4 will focus on the structural, graph-theoretic properties of ILM graphs that we discovered, including distance properties, domination properties, and Hamiltonicity. The novel theoretical results presented in Chapters 3 and 4 were published in [**8**].

In Chapter 5, we will define the Iterated Global Model (IGM). This model is independent of the original graph structure, unlike ILM graphs, and relies on subsets of vertices to generate new graphs. We will prove that graphs generated by a type of IGM called the half-model

exhibits densification, and we will investigate domination and distances in this model. The final chapter recaps the main results and ends with open problems.

CHAPTER 2

Iterated Transitive and Anti-Transitive Models

2.1. Introduction

In the last decade, two deterministic models of complex networks of particular interest to this thesis were introduced: the *Iterated Local Transitivity (ILT)* model and the *Iterated Local Anti-Transitivity (ILAT)* model [**9, 10**]. Consider a social network where friendships have positive edge signs and adversarial relations have negative edge signs. A *triad* is a set of three vertices in a signed network. A triad is said to be *balanced* if the product of the edge signs is positive. Structural balance theory states that these networks seek to balance all triads [**14, 24**]. The ILT and ILAT models were designed with balanced triads in mind.

Throughout this chapter, t will be a non-negative integer. We first define the *Iterated Local Transitivity* model.

DEFINITION 2.1. Given a graph G_0 as input, to form graph G_t in time-step t we *clone* each vertex x, by adding a new vertex x' and adding edges from x' to each vertex in the closed neighbour set of x in the previous iteration of the graph, $N_{G_{t-1}}[x]$. We call x the *parent* of x'.

The idea here is that x' carries the link structure of x. Note that the set of clones forms an independent set. As time progresses, the graph is evolving yet always maintains some

19

structure from the original graph. We next depict the time-steps $t = 1$, 2, 3, and 4 of the

ILT model with the input graph $G_0 = C_4$ from [**9**] in Figure 2.1.
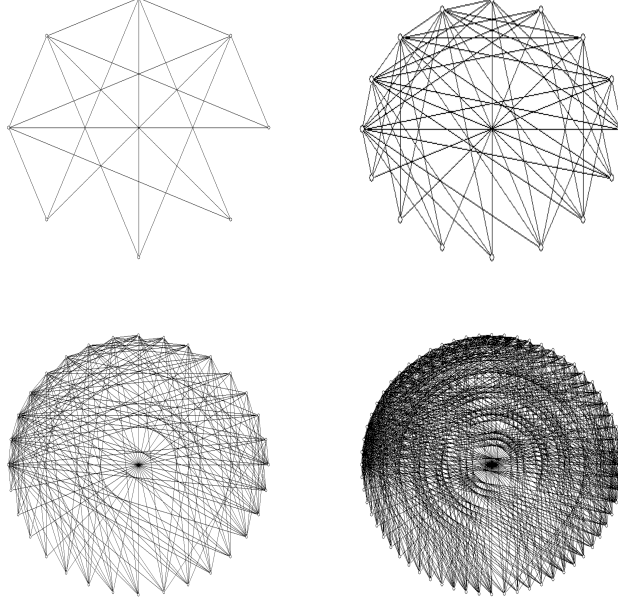
FIGURE 2.1. An example of the first four time-steps of the ILT model, where the initial graph is $G_0 = C_4$, from [**9**].

We next consider the *Iterated Local Anti-Transitivity* model introduced in [**10**]. This model builds on similar ideas from the ILT by replacing the notion of transitivity with *anti-transitivity*; instead of clones, we add anti-clones with edges to each non-neighbour of the parent.

DEFINITION 2.2. Given a graph G_0 as input, to form graph G_t in time-step t we *anti-clone* each vertex x, by adding a new vertex x^* and adding edges from x^* to each vertex not

in the closed neighbour set of x in the previous iteration of the graph, $V(G_{t-1})\backslash N_{G_{t-1}}[x]$. We call x the *parent* of x^*.

Observe that we do not include the edge from x to x^*. For an example, we show the time-steps $t = 1, 2, 3, 4$ of the ILAT model with the input graph $G_0 = C_4$ from [**10**] in Figure 2.2.

We focus on these two models for complex networks, specifically designed for simulating social networks. Both models are applicable to settings where there is some form of duplication of vertices [**9, 10, 21**]. These network models are deterministic and are based on the principles of transitivity and anti-transitivity, as described below in Sections 2.2 and 2.3, respectively.

2.2. Iterated Local Transitivity

We summarize salient results on the ILT model [**9**]. We write $\deg_t(x)$ for the degree of a vertex x at time-step t. The following elementary but important recurrences govern the degrees of vertices in the ILT model. We include this and several other proofs from the original paper for completeness and to give a flavour of the model [**9**].

THEOREM 2.1 ([**9**]). *If G_t is a graph generated by the ILT model at time $t \geq 0$, then we have that*

$$deg_{t+1}(x) = 2deg_t(x) + 1, \tag{2}$$

$$deg_{t+1}(x') = deg_t(x) + 1. \tag{3}$$

21
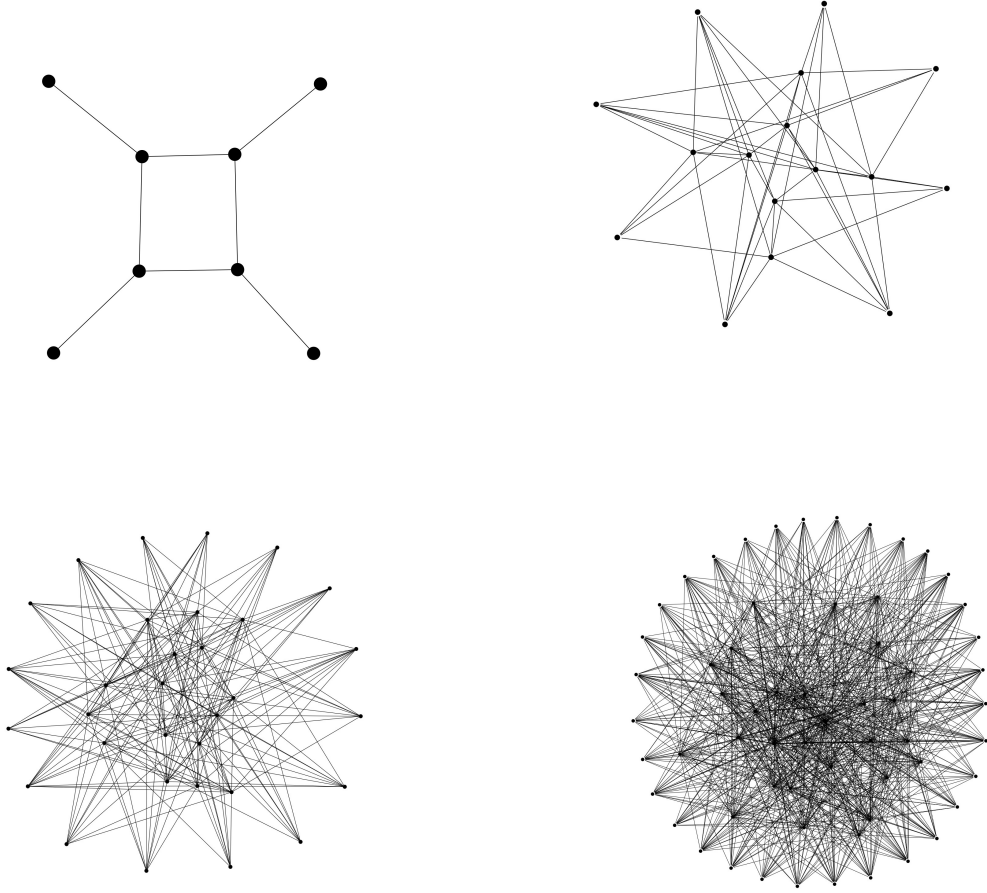
FIGURE 2.2. An example of the first four time-steps of the ILAT model, where the initial graph is $G_0 = C_4$, from [**10**].

PROOF. For equation (2), consider a vertex x that exists in G_t. Each neighbour of x in G_t is still a neighbour of x in G_{t+1}, and each neighbour of x in G_t now has a clone that is adjacent to x in G_{t+1}. There is also an edge between x and its clone, so our final result for the degree of x in G_{t+1} is $\deg_{G_{t+1}}(x) = 2\deg_{G_t}(x) + 1$.

For equation (3), consider a clone, say x', in G_{t+1}. For every neighbour of x, there is one edge from x' to each neighbour. There is also the edge xx'. Thus, we find that $\deg_{G_{t+1}}(x') = \deg_{G_t}(x) + 1$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

We now provide a theorem on the size and order of graphs generated by the model.

THEOREM 2.2 ([9]). *For the ILT graph G_t with $t > 0$, if $n_t = |V(G_t)|$ and $e_t = |E(G_t)|$,* *then*

$$n_t = 2^t n_0 \qquad and \qquad e_t = 3^t(e_0 + n_0) - n_t.$$

PROOF. In each time-step, we are doubling the number of vertices. Hence, we have that $n_t = 2n_{t-1} = 2^t n_0$.

To show that $e_t = 3^t(e_0 + n_0) - n_t$, we use induction on $t \geq 0$. The base case follows by inspection when $t = 0$. We proceed with the inductive step, and assume the equation holds for some t. At time-step $t + 1$, consider how we add new edges to G_t. There is a new edge for each previously formed edge xy in G_t, which gives e_t edges. For each edge xy, we have an edge $x'y$ and an edge $y'x$, which adds $2e_t$ edges to the size of G_{t+1}. Finally, there is a new edge from each vertex x to its clone x', contributing n_t vertices to the size. In summation we have:

$$e_{t+1} = e_t + 2e_t + n_t = 3e_t + n_t.$$

23

Applying the induction hypothesis, we obtain that

$$
\begin{aligned}
e_{t+1} &= 3(3^t(e_0 + n_0) - n_t) + n_t \\
&= 3^{t+1}(e_0 + n_0) - 2n_t = 3^{t+1}(e_0 + n_0) - n_{t+1},
\end{aligned}
$$

and this concludes the proof. □

Recall from Chapter 1 that a graph G_t with e_t edges and n_t vertices satisfies a *densification power law* if there is a constant $a \in (1,2)$ such that e_t is proportional to n_t^a. Such a is called the *densification exponent*. In particular, the average degree grows to infinity with the order of the network, in contrast to the preferential attachment model, which generates graphs with constant average degree [**6**]). Densification power laws were reported in several real-world networks, such as a physics citation graph and the internet graph at the level of autonomous systems [**17**].

The ILT model exhibits a densification power law. Define the *volume* of G_t by

$$
\mathrm{Vol_t} = \sum_{x \in V(G_t)} \deg_t(x) = 2e_t.
$$

Note that the average degree of G_t equals $\mathrm{Vol_t}/n_t$.

THEOREM 2.3 ([**9**]). *For $t > 0$, the average degree of G_t equals*

$$
\left(\frac{3}{2}\right)^t \left(\frac{\mathrm{Vol_0}}{n_0} + 2\right) - 2.
$$

24

Observe that Theorem 2.3 supplies a densification power law with densification exponent

$a = \frac{\log 3}{\log 2} \approx 1.58$.

The proof of Theorem 2.3 follows directly from the following Lemma 2.4, since the average degree of G_t is Vol_t/n_t.

THEOREM 2.4 ([**9**]). *For $t > 0$,*

$$\mathrm{Vol}_t = 3^t \mathrm{Vol}_0 + 2n_0(3^t - 2^t).$$

In particular, $e_t = 3^t(e_0 + n_0) - n_t$.

PROOF. By (2) and (3), we have that

$$
\begin{aligned}
\mathrm{Vol}_{t+1} &= \sum_{x \in V(G_t)} \deg_{t+1}(x) + \sum_{x' \in V(G_{t+1}) \setminus V(G_t)} \deg_{t+1}(x') \\
&= \sum_{x \in V(G_t)} (2\deg_t(x) + 1) + \sum_{x \in V(G_t)} (\deg_t(x) + 1) \\
&= 3\mathrm{Vol}_t + n_{t+1}.
\end{aligned}
\tag{4}
$$

Hence, by (4) for $t > 0$,

$$
\begin{aligned}
\mathrm{Vol}_t &= 3\mathrm{Vol}_{t-1} + n_t \\
&= 3\left(3\mathrm{Vol}_{t-2} + n_{t-1}\right) + 2^t n_0 \\
&= 3^2 \mathrm{Vol}_{t-2} + 3 \cdot 2^{t-1} n_0 + 2^t n_0 \\
&= 3^t \mathrm{Vol}_0 + n_0 \left(\sum_{i=0}^{t-1} 3^i 2^{t-i}\right) \\
&= 3^t \mathrm{Vol}_0 + 2n_0(3^t - 2^t),
\end{aligned}
$$

25

where the third equality follows by summing a geometric series. $\qquad\square$

The next theorem demonstrates that in the ILT model graph, the average distance is bounded by a constant and converges. This suggests that ILT graphs satisfy the small-world property. For the distance between two vertices x and y in G_t, we write $\text{dist}_t(x, y)$.

THEOREM 2.5 ([**9**]). *Let x and y be vertices in G_t with $t > 0$. We then have that*

$$\text{dist}_{t+1}(x', y) = \text{dist}_{t+1}(x, y') = \text{dist}_{t+1}(x, y) = \text{dist}_t(x, y),$$

and

$$\text{dist}_{t+1}(x', y') = \begin{cases} \text{dist}_t(x, y) & \text{if } xy \notin E(G_t), \\ \text{dist}_t(x, y) + 1 = 2 & \text{if } xy \in E(G_t). \end{cases}$$

PROOF. We prove that $\text{dist}_{t+1}(x, y) = \text{dist}_t(x, y)$. The proofs of the other equalities are analogous [**9**], so they are omitted in our exposition. In the ILT model we do not delete any edges, the distance cannot increase after a "cloning" step occurs.

Hence, $\text{dist}_{t+1}(x, y) \leq \text{dist}_t(x, y)$. Suppose for a contradiction that there is a path P' connecting x and y in G_{t+1} with length $k < \text{dist}_t(x, y)$. Hence, P' contains vertices not in G_t. Choose such a P' with the least number of vertices, say $s > 0$, not in G_t. Let z' be a vertex of P' not in G_t, and let the neighbours of z' in P' be u and v. We then have that $z \in V(G_t)$ is adjacent to u and v. Form the path Q' by replacing z' by z. But then Q' has length k and has $s - 1$ many vertices not in G_t, which supplies a contradiction. $\qquad\square$

We now present a theorem on average distance in the ILT model.

THEOREM 2.6 ([**9**]). *For the ILT model graph G_t with $t > 0$, we have that*

$$L(G_t) = 2 \cdot \frac{4^t \left(W(G_0) + (e_0 + n_0) \left(1 - \left(\frac{3}{4}\right)^t \right) \right)}{4^t n_0^2 - 2^t n_0}.$$

PROOF. We derive a recurrence for $W(G_t)$ as follows. To compute $W(G_{t+1})$, there are five cases to consider: distances within G_t, and distances of the forms $\text{dist}_{t+1}(x, y')$, $\text{dist}_{t+1}(x', y)$, $\text{dist}_{t+1}(x, x')$, and $\text{dist}_{t+1}(x', y')$. The first three cases contribute $3W(G_t)$ by Lemma 2.5. The fourth case contributes n_t. The final case contributes $W(G_t) + e_t$ (the term e_t comes from the fact that each edge xy contributes $\text{dist}_t(x, y) + 1$).

We then have that

$$\begin{aligned} W(G_{t+1}) &= 4W(G_t) + e_t + n_t \\[2mm] &= 4W(G_t) + 3^t(e_0 + n_0). \end{aligned}$$

Hence,

$$\begin{aligned} W(G_t) &= 4^t W(G_0) + \sum_{i=0}^{t-1} 4^i \left(3^{t-1-i} \right) (e_0 + n_0) \\[2mm] &= 4^t W(G_0) + 4^t (e_0 + n_0) \left(1 - \left(\frac{3}{4}\right)^t \right), \end{aligned}$$

and this completes the proof. $\qquad\square$

Diameters are constant in the ILT model, and this will be discussed and proven in a more general context in Chapter 4. The ILT model also generates graphs with high clustering coefficients, as we expect in the small-world property. The clustering coefficient of the graph at time-step t generated by the ILT model is estimated and shown to tend to 0 slower than a binomial random graph $\mathbb{G}(n, p)$ with the same average degree, as the following theorem demonstrates as proved in [9].

THEOREM 2.7 ([9]).

$$\Omega\left(\left(\frac{7}{8}\right)^t t^{-2}\right) = C(G_t) = O\left(\left(\frac{7}{8}\right)^t t^2\right).$$

Observe that $C(G_t)$ tends to 0 as $t \to \infty$. If we let $n_t = n$ (so $t \sim \log_2 n$), then this gives that

$$C(G_t) = n^{\log_2(7/8)+o(1)}.$$

In contrast, for a random graph $\mathbb{G}(n, p)$ with average degree $pn = \Theta((3/2)^{\log_2 n}) = \Theta(n^{\log_2(3/2)})$, which is comparable to the above for G_t, has clustering coefficient $p = \Theta(n^{\log_2(3/4)})$ which tends to zero much faster than the previously mentioned $C(G_t)$.

2.3. Iterated Local Anti-Transitivity

We now summarize results and proofs on the ILAT model from the literature [10]. In a graph G with vertex v, the *co-degree* of v, written $\operatorname{codeg}(v)$, is $|V(G)| - \deg(v)$. The following

theorem gives the basic recurrence on degrees in the ILAT model. As in the previous section, we write $\deg_t(x)$ for the degree of x in G_t.

THEOREM 2.8 ([**10**]). *Let G_t be a graph generated by the ILAT model at time-step t. Fix $x \in V(G_t)$ and let $n_t = |V(G_t)|$. We have the following.*

(1) $deg_{t+1}(x) = n_t - 1$.

(2) $deg_{t+1}(x^\star) = n_t - deg_t(x) - 1$.

PROOF. For item (1), note that for all non-neighbours of x, the anti-clone of the non-neighbour in time-step $t + 1$ will have an edge to x. Therefore, neighbours of x and the non-neighbours of x together yield the following formulas:

$$
\begin{aligned}
\deg_{t+1}(x) &= \deg_t(x) + \mathrm{codeg}_t(x) \\
&= \deg_t(x) + (n_t - \deg_t(x) - 1) \\
&= n_t - 1.
\end{aligned}
$$

For item (2), since $x^\star$ is incident to all the non-neighbours of x, we have that

$$
\deg_{t+1}(x^\star) = \mathrm{codeg}_t(x) = n_t - \deg_t(x) - 1,
$$

and this completes the proof. □

The following lemma sets up a recursive formula for the volume, and hence, the number of edges.

THEOREM 2.9 ([**10**]). *For $t \geq 0$ we have that*

$$\mathrm{Vol}_{t+1} = 2n_t^2 - 2n_t - \mathrm{Vol}_t.$$

In particular, we have that $e_{t+1} = n_t^2 - n_t - e_t$.

PROOF. We derive that

$$
\begin{aligned}
\mathrm{Vol}_{t+1} &= \sum_{u \in V(G_{t+1})} \deg_{t+1}(u) \\
&= \sum_{u \in V(G_{t+1})} \deg_t(u) + \sum_{u \in V(G_{t+1})} \deg_t(u') \\
&= \sum_{u \in V(G_t)} n_t - 1 + \sum_{u \in V(G_t)} n_t - \deg_t u - 1 \\
&= 2 \sum_{u \in V(G_t)} n_t - \sum_{u \in V(G_t)} \deg_t(u) - 2 \sum_{u \in V(G_t)} 1 \\
&= 2n_t^2 - 2n_t - \mathrm{Vol}_t,
\end{aligned}
$$

where the third equality follows from Lemma 2.8.

Next, we observe that

$$
\begin{aligned}
e_{t+1} &= \frac{1}{2}\mathrm{Vol}_{t+1} \\
&= n_t^2 - n_t - \frac{1}{2}\mathrm{Vol}_t \\
&= n_t^2 - n_t - e_t.
\end{aligned}
$$

30

where the second equality follows from the previous derivation for the volume. $\qquad\square$

Next, we prove that the ILAT model generates graphs which densify over time; that is, the average degree of the graphs tends to infinity.

THEOREM 2.10 ([**10**]). *In the ILAT model, $\frac{e_t}{n_t} \to \infty$.*

PROOF. By Lemma 2.9, we have that

$$e_{t+1} = (n_0)^2 2^{2t} - n_0 2^t - e_t.$$

Hence, by recursion, we find that

$$
\begin{aligned}
e_{t+1} &= ((n_0)^2 2^{2t} - n_0 2^t) - ((n_0)^2 2^{2(t-1)} - n_0 2^{(t-1)}) + \cdots \\
&= \sum_{i=0}^{t} (-1)^i \left((n_0)^2 2^{2(t-i)} - n_0 2^{(t-i)} \right) \\
&= (n_0)^2 \sum_{i=0}^{t} (-1)^i 2^{2(t-i)} - n_0 \sum_{i=0}^{t} (-1)^i 2^{(t-i)} \\
&= (n_0)^2 S_1 - n_0 S_2,
\end{aligned}
$$

where

$$
\begin{aligned}
S_1 &= \sum_{i=0}^{t} (-1)^i 2^{2(t-i)} \\
&= 2^{2t} - 2^{2t-2} + 2^{2t-4} - 2^{2t-6} + \cdots \\
&= 2^{2t} \left(1 - \frac{1}{2^2} + \frac{1}{2^4} - \frac{1}{2^6} + \cdots \right) \\
&= 2^{2t} \left(\frac{1 - (-\frac{1}{4})^t}{1 + \frac{1}{4}} \right) \\
&= \frac{4}{5} 2^{2t} \left(1 - \left(-\frac{1}{4} \right)^t \right),
\end{aligned}
$$

and where

$$
\begin{aligned}
S_2 &= 2^t - 2^{t-1} + 2^{t-2} - 2^{t-3} + \cdots \\
&= 2^t \left(1 - \frac{1}{2} + \frac{1}{2^2} - \frac{1}{2^3} + \cdots \right) \\
&= 2^t \left(\frac{1 - (-\frac{1}{2})^t}{1 + \frac{1}{2}} \right) \\
&= \frac{2}{3} 2^t \left(1 - \left(-\frac{1}{2} \right)^t \right).
\end{aligned}
$$

Note that we found the value of the infinite alternating sums in the derivations of S_1 and S_2 by using the formula for the sum of a geometric series. Therefore, we have that

$$
\begin{aligned}
e_{t+1} &= (n_0)^2 \frac{4}{5} \, 2^{2t} \left(1 - \left(-\frac{1}{4} \right)^t \right) - n_0 \frac{2}{3} \, 2^t \left(1 - \left(-\frac{1}{2} \right)^t \right) \\
&= (n_0)^2 \frac{1}{5} \, 2^{2(t+1)} \left(1 - (-4)^{-t} \right) - n_0 \frac{1}{3} \, 2^{t+1} \left(1 - (-2)^{-t} \right) \\
&= (n_0)^2 2^{2(t+1)} \left(\frac{1}{5} - \frac{1}{n_0 3(2)^{t+1}} - \frac{1}{5}(-4)^{-t} + \frac{1}{n_0 3(2)^{t+1}}(-2)^{-t} \right) \\
&= (n_0)^2 2^{2(t+1)} \left(\frac{1}{5} - o(1) \right),
\end{aligned}
$$

where the term $o(1)$ denotes terms in t tending to 0 as t tends to ∞. Hence,

$$
e_t = 2^{2t} \left(\frac{(n_0)^2}{5} - o(1) \right).
$$

The latter gives that

$$
\frac{e_t}{n_t} = 2^t \left(\frac{n_0}{5} - o(1) \right),
$$

and this concludes the proof. $\qquad\square$

From this discussion, we can obtain a limiting density as $t \to \infty$. Let D_t be the density of G_t; that is, $D_t = \frac{e_t}{\binom{n_t}{2}}$. Notice that graph density and densification are distinct properties. We finish this chapter by proving that the ILAT model generates graphs that are quite dense.

THEOREM 2.11 ([10]). *As $t \to \infty$, we have that $D_t \to 2/5$.*

33

PROOF. Observe that $D_t = 2\frac{e_t}{n_t(n_t-1)}$. From the proof of Theorem 2.10 and the identity $n_t = 2^t n_0$, we then have that

$$D_t = \frac{2^{2t}(n_0)^2}{n_t(n_t-1)} \left(\frac{1}{5}\right) \left(1 - \left(-\frac{1}{4}^{t-1}\right)\right)(1 - o(1))$$

$$\sim \quad 2/5,$$

and this concludes the proof. $\square$

CHAPTER 3

Iterated Local Model for Social Networks

3.1. Introduction

In many applications of complex networks, both transitivity and anti-transitivity are observed among triads of vertices [**8, 20**]. The ILT and ILAT models restrict their attention considering exactly one of transitivity or anti-transitivity, respectively. We combine the two methods and define the new *Iterated Local Model* (or *ILM*). A *transitive step* is defined as the cloning process from the ILT model, so that in time-step $t > 0$, we add a clone x' for each vertex $x \in V(G_{t-1})$ and define $N_{G_t}(x') = N_{G_{t-1}}[x]$. An *anti-transitive step* is defined as the anti-cloning process from the ILAT model where, in time-step t we add an anti-clone x^* for each vertex $x \in V(G_{t-1})$, and define $N_{G_t}(x^*) = V(G_{t-1}) \backslash N_{G_{t-1}}[x]$.

DEFINITION 3.1. Given an infinite binary sequence $S = \{b_i\}_{i \in \mathbb{N}}$, where $b_i \in \{0, 1\}$, and some input graph G_0. For each time-step $t > 0$, we will define a graph G_t containing G_{t-1} as an induced subgraph. At time-step $t \geq 1$, we consider the value of b_t. If $b_t = 0$, then we take an anti-transitive step, and if $b_t = 1$, then we take a transitive step.

Notice that the constant sequence of 0's will make the ILM equivalent to the ILT model, and a constant sequence of 1's will yield the ILAT model. Hence, the ILM model generalizes

35

both the ILT and ILAT models. Note also that there are uncountably many choices for the binary sequence S.

For a fixed infinite binary sequence S, $\text{ILM}_{t,S}(G)$ denotes the ILM graph generated by the model at time-step t with initial graph G. Figure 3.1 depicts the first four time-steps of ILM for the input graph K_3 and input sequence $S = (0, 1, 0, 1, 0 \ldots)$, the alternating sequence beginning with an anti-transitive step.
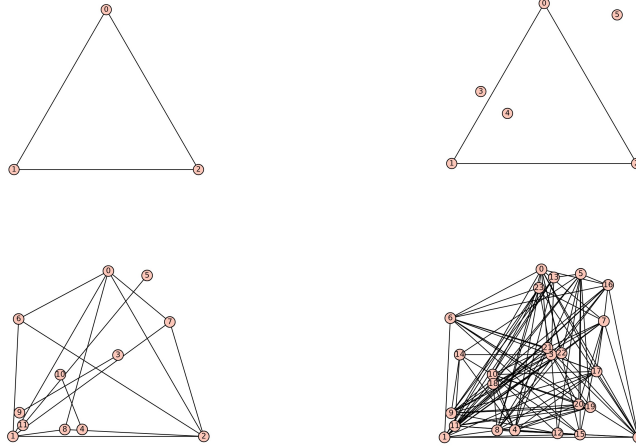
FIGURE 3.1. ILM graphs, where $G_0 = K_3$ for $t = 0, 1, 2, 3$.

To discuss the behaviour of the graph at individual time-steps, we will use the following notation. We denote $\text{LT}(G)$ as the resulting graph after performing a single iteration of a transitive step on the graph G, and analogously we denote $\text{LAT}(G)$ as the resulting graph after performing a single anti-transitive step on the graph G. We will use the term *clones* to refer to all vertices added during a transitive time-step, and the term *anti-clones* to reference all the vertices added during an anti-transitive time-step.

3.2. Complex network properties

To begin this section, we will consider the size of the graphs generated by the Iterated Local Model. To do so, we will first need two theorems regarding the average degrees in both the ILT model and the ILAT model. Recall from Chapter 2 that the volume of a graph G is defined by

$$\mathrm{Vol}(G) = \sum_{v \in V(G)} \deg(v)$$

Hence, the average degree of G is equal to $\mathrm{Vol}(G)/|V(G)|$.

THEOREM 3.1 ([**9**]). *Let G be a graph, and let $G'_t = ILT_t(G)$ and let $G^*_t = ILAT_t(G)$. For all integers $t \geq 1$, the average degree of G'_t equals*

$$\left(\frac{3}{2}\right)^t \left(\frac{\mathrm{Vol}(G'_t)}{n_0} + 2\right) - 2.$$

For all integers $t \geq 1$, the average degree of G^*_t equals

$$2^t \left(\frac{2n_0}{5} - o(1)\right).$$

We now provide an asymptotic formula for the number of edges in an ILM graph, which is the first novel result of the thesis.

THEOREM 3.2. *Let $S = (s_0, s_1, s_2, \dots)$ be a binary sequence with at least one 0 and let $t \geq 0$. Let τ be the first index such that $s_\tau = 0$. For any graph G and all $t \geq \tau$, and with*

$\beta = \beta(t) \le t$ *being the largest index such that* $s_\beta = 0$, *then*

$$|E(ILM_{t,S}(G))| = \Theta\left(2^{t+\beta}\left(\frac{3}{2}\right)^{t-\beta}\right) = \Theta\left(2^\beta\left(\frac{3}{2}\right)^{t-\beta}n_t\right)$$

PROOF. Let $G_t = \text{ILM}_{t,S}(G)$. Consider the case where $s_t = s_\beta = 0$. For all $v \in V(G_{t-1})$, $\deg_t(v) = n_{t-1} - 1$, since for all vertices $u, v \in V(G_{t-1})$, u is a non-neighbour of v if and only if the clone of u is adjacent to v. Therefore, we have that

$$2|E(G_t)| \ge \sum_{v \in V(G_{t-1})} \deg_t(v)$$
$$= \frac{n_t}{2}(n_{t-1} - 1)$$
$$= \frac{n_t(n_t - 2)}{4}.$$

Hence, we find that

$$|E(G_t)| = \Theta(n_t^2) = \Theta(2^{2t}) = \Theta(2^{2\beta}) = \Theta(2^{t+\beta}), \tag{5}$$

and this case is finished.

Consider the case where $s_t = 1$ so $t > \beta$. By (5), we have that $|E(G_\beta)| \ge c2^{2\beta}$ for some constant $c > 0$. By Theorem 3.1, using the graph G_β and the time-step $t - \beta$, we derive that

38

the average degree of G_t is at least

$$\left(\frac{3}{2}\right)^{t-\beta}\left(\frac{2c2^{2\beta}}{2^{\beta}|V(G)|}+2\right)-2\geq\frac{2c}{|V(G)|}2^{\beta}\left(\frac{3}{2}\right)^{t-\beta}.$$

Thus, we have that

$$2|E(G_t)|\geq 2^t\left(\frac{2c}{|V(G)|}2^{\beta}\left(\frac{3}{2}\right)^{t-\beta}\right)$$

$$=\Omega\left(2^{t+\beta}\left(\frac{3}{2}\right)^{t-\beta}\right).$$

Similarly, by (5), we have that there exists a real-valued constant $C>0$ such that $|E(G_{\beta})|\leq C2^{2\beta}$, so by Theorem 3.1, the average degree is at most

$$\left(\frac{3}{2}\right)^{t-\beta}\left(\frac{2C2^{2\beta}}{2^{\beta}|V(G)|}+2\right)-2\leq\left(\frac{3}{2}\right)^{t-\beta}\left(\frac{2C2^{\beta}}{|V(G)|}+2\right),$$

so

$$2|E(G_t)|\leq 2^t\left(\left(\frac{3}{2}\right)^{t-\beta}\left(\frac{2C2^{\beta}}{|V(G)|}+2\right)\right)$$

$$=O\left(2^{t+\beta}\left(\frac{3}{2}\right)^{t-\beta}\right),$$

completing the proof. □

As a corollary to Theorem 3.2, we have that ILM graphs densify for any input sequence.

COROLLARY 3.3. *For any binary sequence S and initial graph G, we have that*

$$\lim_{t \to \infty} \frac{|E(ILM_{t,S}(G))|}{|V(ILM_{t,S}(G))|} = \infty.$$

3.3. Clustering coefficient

Complex networks exhibit high clustering coefficients [**7**]. Recall that clustering coefficients are a measure of local density. In the ILT model, the clustering coefficient tends to 0 as $t \to \infty$, although more slowly than binomial random graphs with the same average degree, as discussed in Chapter 2 [**9**].

For any sequence where there are no infinite sequences of 0s or 1s, we say the sequence has bounded gaps. The ILM for sequences with bounded gaps will have clustering bounded away from 0. We will find it useful to write $c_{t,S,G}(x)$ for $c_{ILM_{t,S}(G)}(x)$, and when S and G are clear from context, we may write $c_t(x)$, consistent with earlier defined notation. The next theorem will bound the change in the clustering coefficient after a single transitive step.

LEMMA 3.4. If G is a graph with minimum degree δ, then

$$C(\mathrm{LT}(G)) \geq \left(\frac{7}{8} - \frac{3}{8\delta}\right) C(G).$$

PROOF. Let $x \in V(G)$ and let x' be its transitive clone in $V(\mathrm{LT}(G))$. Let $c_0(x), c_1(x)$ denote the clustering coefficient of x in G and $\mathrm{LT}(G)$, respectively. Since $N(x') = N[x] \cap V(G)$ and x dominates its neighbourhood, then $c_1(x') \geq c_0(x)$.

We will now give a lower bound for $c_1(x)$. We count the edges in $N(x)$. The induced subgraph $\mathrm{LT}(G)\left[N(x) \cap V(G)\right]$ is the subgraph of $\mathrm{LT}(G)$ induced by the vertex set $N(x) \cap V(G)$. By the definition of clustering, we have that the following number of edges in this subgraph is:

$$c_0 \binom{\deg_G(x)}{2}.$$

We now consider the number of edges between the sets $N(x) \cap V(G)$, the set of old neighbours of x, and $(N(x) \backslash V(G)) \backslash \{x'\}$, the set of new clones adjacent to x other than the clone of x. The number of such edges is:

$$2c_0(x) \binom{\deg_G(x)}{2} + \deg_G(x).$$

Finally, we consider the number of edges within the neighbourhood that are adjacent to x', which is $\deg_G(x)$. To find the clustering coefficient of x, we now must add all of the edges displayed above. The degree of x in $\mathrm{LT}(G)$ is $2\deg_G(x)+1$. We have as the clustering coefficient in $\mathrm{LT}(G)$ satisfies:

$$c_1(x) = \frac{3\binom{\deg_0(x)}{2}c_0(x) + 2\deg_0(x)}{\binom{2\deg_0(x)+1}{2}}$$

$$= \frac{3(\deg_0(x) - 1)c_0(x) + 4}{2(2\deg_0(x) + 1)}$$

$$\geq \left(\frac{3}{4} - \frac{3}{4\deg_0(x)}\right)c_0(x).$$

Since for all $v' \in V(\mathrm{LT}(G)) \setminus V(G)$ we have that $c_1(v') \geq c_0(v)$, we can then find a lower bound for the clustering of the graph $\mathrm{LT}(G)$:

$$C(\mathrm{LT}(G)) = \frac{\sum_{v \in V(\mathrm{LT}(G)) \cap V(G)} c_1(v) + \sum_{v' \in V(\mathrm{LT}(G)) \setminus V(G)} c_1(v')}{2|V(G)|}$$

$$\geq \frac{\sum_{v \in V(G)} \left(\frac{3}{4} - \frac{3}{4\deg_0(v)}\right)c_0(v) + \sum_{v \in V(G)} c_0(v)}{2|V(G)|}$$

$$\geq \frac{\left(\frac{7}{4} - \frac{3}{4\delta(G)}\right)\sum_{v \in V(G)} c_0(v)}{2|V(G)|}$$

$$= \left(\frac{7}{8} - \frac{3}{8\delta(G)}\right)C(G),$$

and this completes the proof. $\qquad \square$

We have the following corollary.

COROLLARY 3.5. *If G is a graph, then*

$$C(ILT_t(G)) = \Omega\left(\left(\frac{7}{8}\right)^t t^{-3/7}\right).$$

PROOF. We define $\delta_t = \delta(\mathrm{ILT}_t(G))$. Since each vertex v' has degree $\deg_{\mathrm{ILT}_{t-1}(G)}(v)+1$ we increase the minimum degree by exactly one, so $\delta_t = \delta_{t-1}+1$ and $\delta_t \geq t$. We may iteratively apply Lemma 3.4 to obtain that

$$C(\mathrm{ILT}_t(G)) \geq C(G)\prod_{i=1}^{t}\left(\frac{7}{8} - \frac{3}{8\delta_i}\right) \geq C(G)\prod_{i=1}^{t}\left(\frac{7}{8} - \frac{3}{8i}\right).$$

Observe that

$$\prod_{i=1}^{t}\left(\frac{7}{8} - \frac{3}{8i}\right) = \left(\frac{7}{8}\right)^t \frac{1}{t!}\prod_{i=1}^{t}\left(i - \frac{3}{7}\right)$$

$$= \left(\frac{7}{8}\right)^t \frac{1}{t!}\prod_{i=1}^{t}\left(t + \frac{4}{7} - i\right)$$

$$= \left(\frac{7}{8}\right)^t \frac{\Gamma(t+4/7)}{t!\Gamma(4/7)}$$

$$= \left(\frac{7}{8}\right)^t \frac{\Gamma(t+4/7)}{t\Gamma(t)\Gamma(4/7)} = \Theta\left(\left(\frac{7}{8}\right)^t t^{-3/7}\right),$$

where Γ is the Gamma function. Thus,

$$C(\mathrm{ILT}_t(G)) = \Omega\left(\left(\frac{7}{8}\right)^t t^{-3/7}\right),$$

which is the required result. $\qquad\square$

43

The following lemma demonstrates that the clustering coefficient is bounded away from zero at time-step t whenever $s_{t-1} = 0$.

LEMMA 3.6. Let G be a graph and $S = (s_0, s_1, s_2, \dots)$ be a binary sequence with bounded gaps between 0's. Let k be a constant such that there is no gap of length k, and let τ be the third index such that $s_\tau = 0$. For all $t \geq \tau$, if $s_t = 0$, then

$$C(\text{ILM}_{t,S}(G)) \geq (1 + o(1)) \frac{1}{2^{2k+4}}.$$

PROOF. Let $G_t = \text{ILM}_{t,S}(G)$ with $t \geq \tau$, and let $\beta_1 < \beta_2 \leq t$ be the two largest time-steps such that $s_{\beta_1} = s_{\beta_2} = 0$. Since we have no gaps of length k then $\beta_1 \geq \beta_2 - k \geq t - 2k$. If $v \in V(G_{\beta_1 - 1})$, then $\deg_{\beta_1}(v) = \frac{n_{\beta_1}}{2} - 1$, where we use the notation $\deg_{\beta_1}(v)$ to be the degree of v in G_{β_1}.

Consider the case where $s_{\beta_1 + 1} = 1$, then

$$\deg_{\beta_1 + 1}(v) = 2\deg_{\beta_1}(v) + 1 = \frac{n_{\beta_1 + 1}}{2} - 1.$$

Since all time-steps between β_1 and β_2 are transitive steps, we have shown inductively that

$$\deg_{\beta_2 - 1}(v) = \frac{n_{\beta_2 - 1}}{2} - 1.$$

44

Thus, v is adjacent to exactly half of the vertices in $V(G_{\beta_2})\backslash V(G_{\beta_2-1})$, since v is not adjacent to itself.

Let $X_{\beta_2} = N(v) \cap (V(G_{\beta_2})\backslash V(G_{\beta_2-1}))$ be the set of clones that are adjacent to v, and $Y_{\beta_2} = (V(G_{\beta_2})\backslash V(G_{\beta_2-1})) \backslash N(v)$ be the set of clones that are not adjacent to v, as depicted in Figure 3.2 where dashed lines represent non-edges. Since v is adjacent to exactly half of the vertices of G_{β_2} and half of the neighbours of v are clones, it follows that $|X_{\beta_2}| = |Y_{\beta_2}| = \frac{n_{\beta_2}}{4}$.
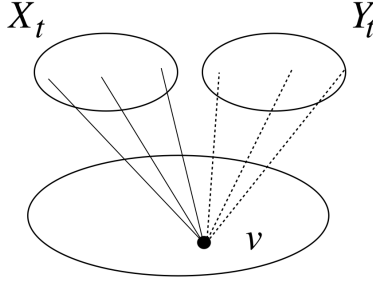
FIGURE 3.2. An illustration of the sets X_{β_2} and Y_{β_2} in Lemma 3.6.

We show inductively that a sequence of transitive time-steps yields two sets with no edges between them and each with cardinality one quarter of the total vertices. Consider the case $s_{\beta_2+1} = 1$, we define X_{β_2+1} and Y_{β_2+1} to be the vertices in X_{β_2} and Y_{β_2}, respectively, along with their new clones in time-step β_2. There are no edges between X_{β_2+1} and Y_{β_2+1} and $|X_{\beta_2+1}| = |Y_{\beta_2+1}| = \frac{n_{\beta_2+1}}{4}$. In addition, $X_{\beta_2+1} \subseteq N(v)$ and $Y_{\beta_2+1} \cap N(v) = \emptyset$. Inductively, we can continue in such a way until we reach time-step $t-1$ with sets X_{t-1} and Y_{t-1}.

We now use the fact that $s_t = 0$. After an anti-transitive step, the vertices in X_{t-1} will be adjacent to every anti-clone of the vertices in Y_{t-1}. We then have that $N(v)$ will contain at least

$$|X_{t-1}| \cdot |Y_{t-1}| = \left(\frac{n_{t-1}}{4}\right)^2 = \left(\frac{n_t}{8}\right)^2 = \frac{n_t^2}{64}.$$

Since v has $\deg(v) = \frac{n_t}{2} - 1$ after an anti-transitive step, then we have that

$$c(v) \geq \frac{n_t^2/64}{\binom{n_t/2-1}{2}} = (1 + o(1))\frac{1}{8}.$$

This holds for all vertices $v \in V(G_{\beta_1 - 1})$. There are $n_{\beta_1-1} \geq n_{t-2k-1} = \frac{n_t}{2^{2k+1}}$ such vertices. Thus,

$$C(G_t) = \frac{(1 + o(1))\left(\frac{1}{8}\right)\left(\frac{n_t}{2^{2k+1}}\right)}{n_t} = (1 + o(1))\left(\frac{1}{2^{2k+4}}\right).$$

This completes the proof. □

We now conclude the section by proving a lower bound for the clustering coefficient for all $\mathrm{ILM}_{t,S}(G)$ given some criteria on S.

THEOREM 3.7. *Let G be a graph, $S = (s_0, s_1, s_2, \dots)$ be a binary sequence with bounded gaps between zeroes, and k be an absolute constant such that there is no gap of length k. If τ*

is the second index such that $s_{tau} = 0$, then for all $t \geq \tau$, the clustering coefficient satisfies:

$$C(ILM_{t,S}(G)) \geq (1 + o(1)) \left(\frac{7}{8}\right)^k \frac{1}{4^{k+2}}.$$

PROOF. Let $G_t = \mathrm{ILM}_{t,S}(G)$ and $\tau \leq \beta \leq t$ denote the largest index such that $s_\beta = 0$. Whenever there are no instances of zeroes between τ and t it is the case that $\beta = \tau$, otherwise $\beta > \tau$. Since we have no gaps of length k, we also have that $\beta \geq t - k$. There must be at least one anti-transitive step prior to β, since $\beta \geq \tau$ and τ is the second instance of an anti-transitive step.

The maximum degree in $G_{\beta-1}$ is $\Delta(G_{\beta-1}) = \frac{n_{\beta-1}}{2} - 1$. Thus, after the anti-transitive step at time-step β, we have the minimum degree is

$$\delta(G_\beta) \geq \frac{n_{\beta-1}}{2} = \frac{n_\beta}{4} \geq \frac{n_t}{2^{k+2}}.$$

The minimum degree of a graph cannot decrease after a transitive step; hence, $\delta(G_{t'}) \geq \frac{n_t}{2^{k+2}}$ for all $t' \geq \beta$. Since thee are only transitive steps between β and t, using Lemma 3.6, we have that

$$C(G_\beta) \geq (1 + o(1)) \left(\frac{1}{2^{2k+4}}\right).$$

By repeated applications of Lemma 3.4 we have that

$$C(G_t) \geq \left(\frac{7}{8} - \frac{3 \cdot 2^{k+2}}{8n_t} \right)^{t-\beta} C(G_{t_1})$$

$$\geq (1 + o(1)) \left(\frac{7}{8} \right)^k \frac{1}{2^{2k+4}},$$

and this completes the proof. □

Determining the clustering coefficient of the ILAT model remains open [**10**]. Theorem 3.7 provides a lower bound of zero when $k = 0$, which is the case in the ILAT model.

3.4. Spectral expansion

In this section, we provide background on spectral graph theory [**12**].

DEFINITION 3.2. For any graph G, the *adjacency matrix*, written $A = (a_{ij})_{n \times n}$, is defined by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i v_j \in E(G), \\ 0 & \text{otherwise.} \end{cases}$$

DEFINITION 3.3. For any graph G, the *diagonal degree matrix*, written $D = (d_{ij})_{n \times n}$, is defined by

$$
d_{ij} =
\begin{cases}
\deg(v_i) & \text{if } i = j, \\[2ex]
0 & \text{otherwise.}
\end{cases}
$$

For a graph G and sets of vertices $X, Y \subseteq V(G)$, define $E(X, Y)$ to be the set of edges in G with one endpoint in X and the other in Y. For simplicity, we write $E(X) = E(X, X)$. Let A denote the adjacency matrix and D denote the diagonal degree matrix of a graph G. Since D is a diagonal matrix, we may find $D^{-1/2}$ by taking the reciprocal of the square root of each entry along the diagonal. The *normalized Laplacian* of G is

$$
\mathcal{L} = I - D^{-1/2} A D^{-1/2}.
$$

Let $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1} \leq 2$ denote the eigenvalues of $\mathcal{L}$. The *spectral gap* of the normalized Laplacian is defined as

$$
\lambda = \max\{|\lambda_1 - 1|, |\lambda_{n-1} - 1|\}.
$$

For sets of vertices X and Y, we use the notation $\mathrm{Vol}(X) = \sum_{v \in X} \deg(v)$ for the volume of X, $\overline{X} = V \setminus X$ for the complement of X, and, $e(X, Y)$ for the number of edges with one end in each of X and Y. It is not necessary for $X \cap Y$ to be empty. In particular, $e(X, X) = 2|E(X)|$. We next state the expander mixing lemma for the normalized Laplacian [12].

LEMMA 3.8 ([**12**]). *If G is a graph with spectral gap λ, then for all sets $X \subseteq V(G)$,*

$$\left| e(X, X) - \frac{(\mathrm{Vol}(X))^2}{\mathrm{Vol}(G)} \right| \leq \lambda \frac{\mathrm{Vol}(X)\mathrm{Vol}(\overline{X})}{\mathrm{Vol}(G)}.$$

THEOREM 3.9. *Let G be a graph and $S = (s_0, s_1, s_2 \ldots)$ be a binary sequence. For all $t \geq 1$, we have that*

$$\lambda_t \geq \frac{1}{4} - o(1),$$

where λ_t is the spectral gap of $ILM_{t,S}(G)$.

PROOF. Let $G_t = \mathrm{ILM}_{t,S}(G)$. First, consider the case where $s_{t-1} = 1$, so that $G_t = \mathrm{LT}(G_{t-1})$. Let $X = V(G_t) \backslash V(G_{t-1})$ be the set of cloned vertices added to G_{t-1} to form G_t. Since X is an independent set, we note that $e(X, X) = 0$. We derive the following using Theorem [2.2]

$$\mathrm{Vol}(G_t) = 6e_{t-1} + 2n_{t-1},$$

$$\mathrm{Vol}(X) = 2e_{t-1} + n_{t-1},$$

$$\mathrm{Vol}(\overline{X}) = \mathrm{Vol}(G_t) - \mathrm{Vol}(X) = 4e_{t-1} + n_{t-1}.$$

Hence, by Lemma 3.8, we have that

$$\lambda_t \geq \frac{(\mathrm{Vol}(X))^2}{\mathrm{Vol}(G_t)} \cdot \frac{\mathrm{Vol}(G_t)}{\mathrm{Vol}(X)\mathrm{Vol}(\overline{X})}$$

$$= \frac{\mathrm{Vol}(X)}{\mathrm{Vol}(\overline{X})}$$

$$= \frac{2e_{t-1} + n_{t-1}}{4e_{t-1} + n_{t-1}} > 1/2.$$

Second, consider the case where $s_{t-1} = 0$, so that $G_t = \mathrm{LAT}(G_{t-1})$. Let $X = V(G_t)\backslash V(G_{t-1})$ be the set of anti-clones added to G_t to form G_{t+1}. We derive that

$$\mathrm{Vol}(G_t) = 2n_{t-1}^2 - 2e_{t-1} - 2n_{t-1},$$

$$\mathrm{Vol}(X) = n_{t-1}^2 - 2e_{t-1} - n_{t-1}, \quad \text{and}$$

$$\mathrm{Vol}(\overline{X}) = \mathrm{Vol}(G_t) - \mathrm{Vol}(X) = n_{t-1}^2 - n_{t-1}.$$

Hence, by Lemma 3.8, we have that

$$\lambda_t \geq \frac{\mathrm{Vol}(X)}{\mathrm{Vol}(\overline{X})}$$

$$= \frac{n_{t-1}^2 - 2e_{t-1} - n_{t-1}}{n_{t-1}^2 - n_{t-1}}$$

$$= 1 - \frac{2e_{t-1}}{n_{t-1}^2 - n_{t-1}} = \frac{1}{4} - o(1),$$

where the last equality follows since if $t \geq 2$, G_{t-1} has an independent set of cardinality $n_{t-1}/2$. In particular, we have that

$$2e_{t-1} \leq 2\binom{n_{t-1}}{2} - 2\binom{n_{t-1}/2}{2} = \frac{3}{4}n_{t-1}^2 - \frac{1}{2}n_{t-1},$$

and this concludes the proof. $\qquad\square$

CHAPTER 4

Structural properties of the Iterated Local Model

4.1. Introduction

In this chapter, we consider the graph-theoretic structure of the graphs generated by the Iterated Local Model. These results focus on graph parameters including diameter, chromatic number, and domination number. We continue the chapter by discussing the conditions under which the model will yield graphs that are Hamiltonian. We finish the chapter by providing an interesting result regarding induced subgraphs of the model: we prove that every finite graph is contained as an induced subgraph of the model after some time-step for every initial graph or sequence.

4.2. Graph parameters

In this section, we will discuss various graph parameters in the ILM model.

4.2.1. Chromatic Number. Before we prove the theorem on the chromatic number of the ILM model, we must first introduce a few lemmas.

LEMMA 4.1. Let F be an induced subgraph of a graph H. If for every vertex $v \in V(F)$, there exists $u \in V(H) \backslash V(F)$ with $N[v] \cap V(F) \subseteq N(u)$, then $\chi(H) > \chi(F)$.

PROOF. See Figure 4.1.
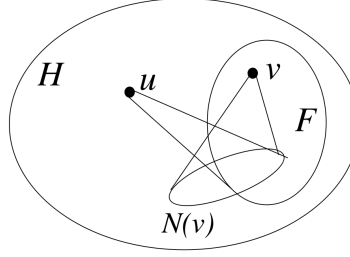
FIGURE 4.1. A diagram of the neighbourhood conditions for Lemma 4.2.

We prove this claim by contradiction. Assume $\chi(H) = \chi(F)$. The proper colouring of H also induces a proper colouring of F. Thus, there must be some vertex $v \in V(F)$ such that $N[v] \cap V(F)$ witnesses every colour. In H, the neighbour $u \in V(H) \backslash V(F)$ of vertex $v \in V(F)$ is adjacent to every vertex in $N[v] \cap V(F)$, and thus every colour in the set of $\chi(H)$ colouurs. Hence, there is no available colour to colour u. This is a contradiction. Thus, it must be the case that $\chi(H) > \chi(F)$.

$\square$

LEMMA 4.2. If G is a connected graph, then $\chi(\mathrm{LT}(G)) = \chi(G) + 1$. If the radius of G is at least 3, then $\chi(\mathrm{LAT}(G)) = \chi(G) + 1$.

PROOF. In both $\mathrm{LT}(G)$ and $\mathrm{LAT}(G)$, we have that the vertex set of G_t in each case can be partitioned into two sets: an independent set that consists of all the new clones, and the set of all vertices in G_{t-1}. Additionally, In both $\mathrm{LT}(G)$ and $\mathrm{LAT}(G)$, $G_t[G_{t-1}] = G_{t-1}$, since all the edges in $E(G_t) \backslash E(G_{t-1})$ contain an endpoint in $V(G_t) \backslash V(G_{t-1})$. In either case, since we may colour all new vertices a single colour. Thus, we have that

$$\chi(G) \leq \chi\left(\mathrm{LT}(G)\right) \leq \chi(G) + 1 \qquad \text{and} \qquad \chi(G) \leq \chi\left(\mathrm{LAT}(G)\right) \leq \chi(G) + 1.$$

For the LT case, we use lemma 4.1, where G and $\mathrm{LT}(G)$ are the graphs F, H respectively as stated in the lemma.

We continue the proof of the lemma for $\mathrm{LAT}(G)$. Let G be a graph of radius at least three. For every $v \in V(G)$ there exists a vertex $u \in V(G)$ with $\mathrm{dist}(u,v) \geq 3$. Consider the anti-clone of u, $u^* \in V\left(\mathrm{LAT}(G)\right) \backslash V(G)$. Every neighbour in G of v is not adjacent to u or else the distance from v to u would be only two. Hence, each neighbour in G of v will be adjacent to u^*. See Figure 4.2. This is the same property as Lemma 4.1, since $\mathrm{LAT}(G)$ has G as an induced subgraph. Thus, $\chi\left(\mathrm{LAT}(G)\right) > \chi(G)$ whenever G has radius three.
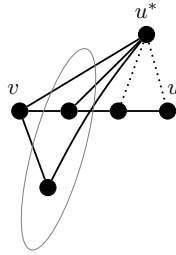

FIGURE 4.2. A diagram of the anti-transitive case for Lemma 4.2.

Further, in each case, we are adding an independent set. Thus, the chromatic number at each time-step can increase by at most one colour, so $\chi\left(\mathrm{LAT}(G)\right) \leq \chi(G) + 1$ and

$\chi\left(\mathrm{LT}(G)\right) \leq \chi(G) + 1$. The upper and lower bounds together yield the result as stated in the lemma. $\qquad\square$

The following lemma will give a lower bound on the radius of graphs after an anti-transitive step. This will help us towards our proof of the chromatic number for all graphs generated by the Iterated Local Model. Subsequently, we will prove an additional lemma relating the radius and diameter after a transitive step.

LEMMA 4.3. For any graph G, the radius of $\mathrm{LAT}(G)$ is at least three.

PROOF. Consider a vertex $x \in V(G)$ and its anti-clone $x^* \in V\left(\mathrm{LAT}(G)\right) \backslash V(G)$. We note that $N[x] \cap N[x^*] = \emptyset$. Therefore, for every vertex x, there exists some vertex x^* that has dist $(x, x^*) \geq 3$. This completes the proof. $\qquad\square$

We next prove the following.

LEMMA 4.4. If G is a graph with radius at least 3, then the radius of $\mathrm{LT}(G)$ is at least 3. Further, whenever $\mathrm{diam}(G) \neq 1$, $\mathrm{diam}(G) = \mathrm{diam}\left(\mathrm{LT}(G)\right)$.

PROOF. Consider $u, v \in V(G)$. We first prove the following claim.

Claim: $\mathrm{dist}_G(u, v) = \mathrm{dist}_{\mathrm{LT}(G)}(u, v)$

Consider any vertices $u, v \in V(G)$. We must show that we cannot find a shorter path using the new vertices in $\mathrm{LT}(G)$. For any vertex $w \in V(G)$ and its transitive clone $w' \in \mathrm{LT}(G)$ with $N(w') = N[w] \cap V(G)$, there is no uv-shortest path that contains both w and

56

w'. Thus, there exists a shortest path between u, v in $\mathrm{LT}(G)$ that uses only vertices from $V(G)$. This proves the claim.

Further, for any uv-shortest path in $V(G)$, we can consider any interior vertex on this path, and replace it along the path with its respective clone $w' \in V(\mathrm{LT}(G))$ and obtain a new uv-path in $\mathrm{LT}(G)$ of the same length.

Consider the path from u' to v'. Whenever u, v are not adjacent, we may use the uv-shortest path containing only $V(G)$ vertices and have that $\mathrm{dist}_{\mathrm{LT}(G)}(u', v') = \mathrm{dist}_G(u, v)$. However, whenever u, v are adjacent, we must first move from u' to u before arriving at v'. Thus, if $\mathrm{dist}_G(u, v) = 1$ then $\mathrm{dist}_{\mathrm{LT}(G)}(u', v') = 2$. Whenever G has diameter at least two, $\mathrm{diam}(G) = \mathrm{diam}\,(\mathrm{LT}(G))$. $\qquad\square$

With the two preceding lemmas, we are ready to complete the proof of the chromatic number for the Iterated Local Model.

THEOREM 4.5. *If G is a connected graph and $S = (s_0, s_1, s_2, \ldots)$ is a binary sequence, then the chromatic number satisfies*

$$\chi(G) + t - 1 \leq \chi\left(ILM_{t,S}(G)\right) \leq \chi(G) + t,$$

for all $t \geq 0$.

PROOF. For the upper bound, we may consider colouring the independent set at each time-step with a new colour. This will always achieve a proper colouring with $\chi(G) + t$ many colours at time-step $t \geq 0$.

For the lower bound, we first consider a sequence of constant ones, in which case by repeated application of Lemma 4.2, we have that

$$\chi \left(\mathrm{ILT}_t(G) \right) = \chi(G) + t$$

for all $t \geq 0$.

Consider now the first index, τ such that $s_\tau = 0$. We then have that $\mathrm{ILM}_{\tau-1,S}(G) = \mathrm{ILT}_{\tau-1}(G)$. Thus, we have that

$$\chi \left(\mathrm{ILM}_{\tau-1,S}(G) \right) = \chi \left(\mathrm{ILT}_{\tau-1}(G) \right) = \chi(G) + (\tau - 1).$$

Since $\mathrm{ILM}_{\tau-1,S}(G)$ is an induced subgraph of $\mathrm{ILM}_{\tau,S}(G)$, it must be the case that

$$\chi \left(\mathrm{ILM}_{\tau,S}(G) \right) \geq \chi \left(\mathrm{ILM}_{\tau-1,S}(G) \right) = \chi(G) + \tau - 1.$$

For all $t \geq \tau$, $\mathrm{ILM}_{t,S}(G)$ has radius at least 3 by Lemma 4.3, we may then use Lemma 4.2 to determine the following bound for any $t \geq \tau$,

$$\chi \left(\mathrm{ILM}_{t,S}(G) \right) \geq \chi \left(\mathrm{ILM}_{t-1,S}(G) \right) + 1.$$

Since we may apply Lemma 4.2 for each time-step between τ and t, we then have that

$$\chi \left(\mathrm{ILM}_{t,S}(G) \right) \geq \chi(G) + (t - \tau) + \tau - 1 = \chi(G) + t - 1.$$

This completes the proof. □

4.2.2. Domination. We next focus on the domination number of ILM graphs.

THEOREM 4.6. *Let G be a graph and $S = (s_0, s_1, s_2, \dots)$ be a binary sequence with at least two bits equal to 0. If τ is the second index such that $s_\tau = 0$, then for all $t \geq \tau + 1$,*

$$\gamma \left(ILM_{t,S}(G) \right) \leq 3.$$

PROOF. Let $G_t = \mathrm{ILM}_{t,S}(G)$ for any specified t and $\beta_1 < \beta_2 \leq t$ be the largest and second largest indices such that $s_{\beta_1} = s_{\beta_2} = 0$. It may be the case that $\beta_2 = \tau$, but this is not necessary.

Let $v \in V(G_{\beta_1})$. Since $s_{\beta_1} = 0$, $G_{\beta_1+1} = \mathrm{LAT}(G_{\beta_1})$. Let $v^* \in V(G_{\beta_1+1})$ be the anti-clone of v. Further, let $v^{**} \in V(G_{\beta_2+1})$ be the anti-clone of v^*.

Claim: The set $\{v, v^*, v^{**}\}$ is a dominating set of G_{β_2+1}.

Any vertex $u \in V(G_{\beta_2})$ is adjacent to one of either v^* or v^{**} by definition of anti-clone. We must now consider new vertices in $V(G_{\beta_2+1})\backslash V(G_{\beta_2})$. These vertices form an independent set, and so none are adjacent to v^{**}. Also, $N_{\beta_1+1}[v] \cap N_{\beta_1+1}[v^*] = \emptyset$, by the definition of anti-clone. Since we only witness transitive steps between time-step β_1 and β_2, we also have that $N_{\beta_2}[v] \cap N_{\beta_2}[v^*] = \emptyset$. Thus, every vertex $y \in V(G_{\beta_2})$ is not adjacent to at least one of v, v^*. We then have that the anti-clone of y, $y^* \in V(G_{\beta_2+1})\backslash V(G_{\beta_2})$ must be adjacent to at least one of v, v^*. Therefore, $\{v, v^*, v^{**}\}$ is a dominating set of G_{β_2+1}.

To complete the proof, we notice that transitive steps will preserve domination. A dominating set in G_t will continue to dominate in $G_{t+1} = \mathrm{LT}(G_t)$ since the clones of the vertices outside of the dominating set are also adjacent to the dominating set. Therefore, the domination number of $\mathrm{ILM}_{t,S}(G)$ for any sequence S and any graph G is bounded above by three, when $t \geq \beta_2 + 1$. □

For any graph, an anti-transitive step cannot yield a dominating vertex since the anti-clone is never adjacent to the parent. Theorem 4.6 bounds the domination number from above. We cannot have $\gamma\left(\mathrm{ILM}_{t,S}(G)\right) = 1$ whenever S has at least one zero. Therefore, $\mathrm{ILM}_{t,S}(G)$ can only witness a domination of two or three whenever $\mathrm{ILM}_{t,S}(G) \neq \mathrm{ILT}_t(G)$. We continue in the next theorem to determine exactly when we have domination number two for these graphs. We first present a lemma regarding vertex partitions.

LEMMA 4.7. Let G be a graph. If G contains a pair of vertices whose closed neighbourhoods partition the vertex set, then the same pair of closed neighbourhoods also partition the vertex set in $\mathrm{LAT}(G)$ and $\mathrm{LT}(G)$. Furthermore, if $\mathrm{LT}(G)$ contains a pair of vertices whose closed neighbourhoods partition the vertex set, then G also has such a pair.

PROOF. Let $u, v \in V(G)$ be a pair of vertices whose closed neighbourhoods partition the vertex set. Consider first, an anti-transitive step. We then have that every anti-clone of the vertices in $N[u] \cap V(G)$ is adjacent to v, and every anti-clone of the vertices in $N[v] \cap V(G)$ is adjacent to u. Thus, $V(\mathrm{LT}(G)) = N[u] \cup N[v]$. Also, since $N[u], N[v]$ partition the vertex set of G, and with the properties of the anti-clone, we have that $N[u] \cap N[v] = \emptyset$ as required.

60

Consider the transitive case. We claim that $N[u]$ and $N[v]$ will partition the vertex set of LT(V). Every transitive clone of $N[u] \cap V(G)$ is adjacent to u and every transitive clone of $N[v] \cap V(G)$ is adjacent to v. Hence, we have that $N[u] \cup N[v] = V(\mathrm{LT}(G))$. Additionally, no clone of $N[u] \cap V(G)$ is adjacent to $N[v] \cap V(G)$ and no clone of $N[v] \cap V(G)$ is adjacent to $N[u] \cap V(G)$. Thus, $N[u] \cap N[v] = \emptyset$, as required.

Moving on to the second part of the lemma, let $x, y \in V(\mathrm{LT}(G))$ be two vertices that partition the vertex set of LT(G). If both $x, y \in V(G)$ then they must partition the vertex set of G since G is an induced subgraph of LT(G). If not, then there exists some vertex $w \in V(G)$ such that, without loss of generality, $w' = x$.

Case 1: Suppose that there also exists some vertex $z \in V(G)$ such that $z' = y$.

Since x and y have neighbourhoods that partition the vertex set and all clones form an independent set, it must be the case that x, y are the only two clones. Therefore, either $G = K_2$ or $G = \overline{K_2}$ (that is, a disjoint pair of edges). In both cases we consider taking a transitive step. If $G = K_2$, then we consider the graph of G shown in Figure 4.3, and notice that there does not exist a vertex partition in LT(G), which is a contradiction. If $G = \overline{K_2}$, then we have that w, z are the vertices whose neighbourhoods partition the vertex set of G.

Case 2: Suppose now that $x \in V(\mathrm{LT}(G)) \backslash V(G)$ and $w' = x$ for some $w \in V(G)$ and that $y \in V(G)$, without loss of generality.

Since $V(\mathrm{LT}(G)) \backslash V(G)$ is an independent set, y must be adjacent to all vertices in $(V(\mathrm{LT}(G)) \backslash V(G)) \backslash \{x\}$. Hence, y is adjacent to $V(G) \backslash \{w\}$. Since $N[x] \cap N[y] = \emptyset$, w
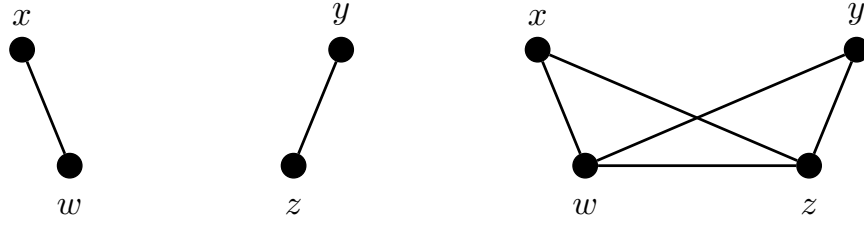
FIGURE 4.3. An illustration of LT(G) in Case 1 of Lemma 4.7 when $G = \overline{K_2}$ or $G = K_2$, respectively.

must then be an isolated vertex in G, and $N[w]$ and $N[y]$ partition the vertex set of G. This concludes the proof. □

The following proof characterizes exactly which graphs, G, and sequences S that will generate ILM graphs with domination two. For this proof we use the notation $G_t = \mathrm{ILM}_{t,S}(G)$.

THEOREM 4.8. *Let G be a graph and let $S = (s_0, s_1, s_2 \dots)$ be any binary sequence that contains at least one bit equal to 0. If τ is the first index such that $s_\tau = 0$, then for all $t \geq \tau + 1$, $\gamma(ILM_{t,S}(G)) = 2$ if and only if one of the following statements holds.*

(1) *The graph G has a pair of vertices whose closed neighbourhoods partition the vertex set.*

(2) *The graph G contains an isolated vertex and $\tau = 0$.*

(3) *The graph G contains a dominating vertex, $s_\tau = s_{\tau+1} = 0$, and $t \geq \tau + 2$.*

PROOF. We begin with the reverse direction. In each case listed by the items of the theorem, we calculate the domination number.

(1) If G has such a pair, then so does G_t for all $t \geq 0$ by Lemma 4.7. Thus, $\gamma(G_t) = 2$.

(2) Let v be the isolated vertex and v^* its anti-clone in G_1. These vertices have neighbourhoods that partition the vertex set since $N(v^*) = V(G)\backslash\{v\}$.

(3) Since τ is the first index of zero, $G_\tau = \text{ILT}_\tau(G)$ which must also have a dominating vertex, say v. After the anti-transitive step, v^* will be an isolated vertex. We consider $G_{\tau+1}$ to be a graph H and the first entry in its sequence is considered to be $s_{\tau+1} = 0$, and we may apply case (2), the result holds.

We now proceed to prove the forward direction. Consider the maximum degree at time-step $\tau + 1$, where we use n_t to denote the number of vertices of G_t at time-step t for any time-step $t \geq 0$. Hence, we have the maximum degree is

$$\Delta(G_{\tau+1}) = \frac{n_{\tau+1}}{2} - 1.$$

When performing an anti-transitive step, for any $v \in V(G_\tau)$ we have that

$$\deg_{G_{\tau+1}}(v) = \deg_{G_\tau}(v) + \left(n_\tau - \deg_{G_\tau}(v)\right) = n_\tau - 1 = \frac{n_{\tau+1}}{2} - 1.$$

If $s_{\tau+1} = 1$, then a vertex in $G_{\tau+1}$ with $\Delta(G_{\tau+1})$ neighbours will have degree

$$2\Delta(G_{\tau+1}) + 1 = n_{\tau+1} - 1 = \frac{n_{\tau+2}}{2} - 1.$$

If $s_{\tau+1} = 0$, then a vertex in $G_{\tau+1}$ with $\Delta(G_{\tau+1})$ neighbours will have degree

63

$$\Delta(G_{\tau+1}) + (n_{\tau+1} - 1 - \Delta(G_{\tau+1})) = \frac{n_{\tau+2}}{2} - 1.$$

Thus, we have proven inductively for all t that $\Delta(G_t) = \frac{n_t}{2} - 1$. With this exact value of the maximum degree, it must be the case that if $\gamma(G_t) = 2$, then G_t must have a pair of vertices whose closed neighbourhoods partition the vertex set. We now must show that the only conditions that yield such a partition are those in the statement of the theorem.

From Lemma 4.7 if G has a pair of vertices whose closed neighbourhoods partition the vertex set of G, then G_t must also have such a pair. This is condition (1).

We next consider G_θ with $\theta \leq t$ as the first time-step that witnesses a pair of vertices that partition the vertex set of G_θ, and let u, v be these vertices. From Lemma 4.7, we know that $s_{\theta-1} = 0$, otherwise u, v would partition $G_{\theta-1}$ which contradictions G_θ being the first graph with such a partition. Furthermore, u, v cannot both be in $V(G_{\theta-1})$ by the same argument. Thus, at least one of u or v is an anti-clone.

Claim: At most one of u or v is an anti-clone at time-step θ.

Analogously to the proof of Lemma 4.7, if v is also an anti-clone, then u and v must be the only two anti-clones. Since $\theta > 0$, the initial graph G must be K_1 and $\theta = 2$. Since $s_\theta = 0$, we only have two possible cases for G_θ: where $s_1 = 1$ or $s_1 = 0$, as illustrated in Figure 4.4.

When $s_1 = 1$, we do not have a partition in G_θ, this is a contradiction. When $s_1 = 0$, $G_{\theta-1}$ also has a pair of vertices that partition the vertex set, which contradicts that G_θ is
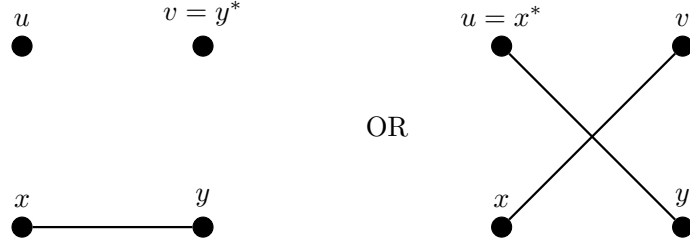
FIGURE 4.4. An illustration of the two possible cases in Theorem 4.8.

the first instance with that property. By contradiction, u, v cannot both be anti-clones, and this concludes the proof of the claim.

Without loss of generality, let u be the vertex in $V(G_\theta) \backslash V(G_{\theta-1})$. and $v \in V(G_{\theta-1})$. Let v^* be the anti-clone of v in G_θ. For the sake of contradiction, suppose that $v^* \neq u$, then $v^* \notin N[v]$ by definition of anti-clone, and since all anti-clones form and independent set, $v^* \notin N[u]$. Thus, v^* is not dominated in G_θ, a contradiction.

Thus, v must be adjacent to all anti-clones except for $u = v^*$ which implies v must be an isolated vertex in $G_{\theta-1}$. If $G_{\theta-1} = G$ then G has an isolated vertex, and we are in condition (2), so we are done. Thus, assume $\theta - 1 \geq 1$. We then find that $G_{\theta-1}$ has an isolated vertex. Hence, $s_{\theta-1} = 0$ since transitive steps cannot create an isolated vertex.

Claim: $G_{\theta-2} = K_1$ or $v \in V(G_{\theta-1}) \backslash V(G_{\theta-2})$.

Suppose not, then $v \in V(G_{\theta-2})$. Since v is isolated in $G_{\theta-1}$, we have that v must also be isolated in $G_{\theta-2}$. Since $G_{\theta-2} \neq K_1$, then there exists a non-neighbour of v in $G_{\theta-2}$. However,

this non-neighbour will have an anti-clone that will then be adjacent to v in $G_{\theta-1}$, which contradicts v being isolated in $G_{\theta-1}$. This ends the proof of the claim.

Additionally, if $G_{\theta-2} = K_1 = G$ and $\theta - 2 = \tau = 0$ and K_1 has a dominating vertex, $s_0 = s_1 = 0$, so we are in condition (3). Therefore, we must now consider the case where $v \in V(G_{\theta-1}) \backslash V(G_{\theta_2})$ and $G_{\theta-2} \neq K_1$. Let $x \in V(G_{\theta-2})$ be the vertex so that $x^* = v$. Since v is isolated, x must then be a dominating vertex of $G_{\theta-2}$. Dominating vertices can only occur from transitive steps, and so $s_i = 1$ for all $i < \theta - 2$. Therefore, $\theta - 2 = \tau$ and $s_\tau = s_{\tau+1} = 0$ and $t \geq \tau + 2$. Since $G_{\tau-2}$ has a dominating vertex, G must also have a dominating vertex, so we are in condition (3). □

4.2.3. Diameter. We continue our discussion of the domination number and distances within graphs generated by the Iterated Local Model. We now work towards discussing the diameter of these graphs in given cases.

LEMMA 4.9. For any graph G, The graph $\mathrm{LAT}(G)$ is disconnected if and only if G has a dominating vertex, or is the disjoint union of two complete graphs.

PROOF. We begin with the forward direction by contraposition. Assume then that G has no dominating vertex and G is not the disjoint union of cliques. Let $A_1, A_2, \ldots, A_\ell$ be the components of G, and let $A_i^* \subseteq V(\mathrm{LAT}(G)) \backslash V(G)$ denote the anti-clones of the vertices in A_i for all $1 \leq i \leq \ell$. Every vertex in A_i is adjacent to every vertex in A_j^* whenever $i \neq j$. Whenever $\ell \geq 3$, $\mathrm{LAT}(G)$ is connected, as in Figure 4.5, and we are done.
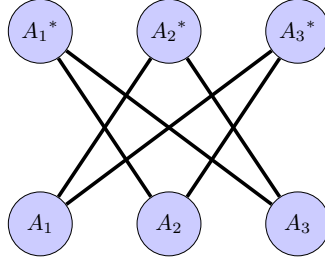
FIGURE 4.5. An illustration of the case where G contains at least three connected components after one anti-transitive step in Lemma 4.9.

Consider the case when $\ell = 2$. All vertices in $A_1 \cup A_2^*$ are connected and all vertices in $A_2 \cup A_1^*$ are connected. Since G is not the disjoint union of two cliques, there exist $u, v \in A_1$, without loss of generality, such that $uv \notin E(G)$. Therefore, $u \in A_1$ is adjacent to $v^* \in A_1^*$, the anti-clone of v. Thus, $\mathrm{LAT}(G)$ is connected.

Consider now the case when $\ell = 1$. Since G does not have a dominating vertex, then for any vertex $u \in V(G)$ there exists a vertex $v \in V(G)$ such that $uv \notin E(G)$. Hence, the anti-clone of u, u^* is adjacent to v. Thus, every anti-clone has some neighbour, so no anti-clone is isolated. Thus, $\mathrm{LAT}(G)$ is connected. This concludes the proof of the forward direction by contraposition.

To prove the reverse direction, consider the case when G has a dominating vertex. We then have that $\mathrm{LAT}(G)$ has an isolated vertex and is disconnected. Next, consider the case when G is the disjoint union of two cliques call them A_1 and A_2. We then have that there are no edges from A_i^* to A_i for $i \in \{1, 2\}$. Thus, $\mathrm{LAT}(G)$ is disconnected. This completes the proof of the reverse direction. $\square$

We next consider the following lemma.

LEMMA 4.10. For any graph, G, if $\gamma(G) \geq 3$, then diam $(\mathrm{LAT}(G)) \leq 3$.

PROOF. Let x and y be in $V(G)$. Since $\gamma(G) \geq 3$, there exists a vertex z that is not adjacent to either x or y, regardless of whether $xy \in E(G)$. If z^* is the anti-clone of z in $\mathrm{LAT}(G)$, then $\{x, z^*, y\}$ is a path in $\mathrm{LAT}(G)$, so $\mathrm{dist}_{\mathrm{LAT}(G)}(x, y) \leq 2$.

Analogously, $\mathrm{dist}_{\mathrm{LAT}(G)}(x, z) = 2$ and $\mathrm{dist}_{\mathrm{LAT}(G)}(y, z) = 2$, since there is always a third vertex not adjacent to the other two, whose anti-clone will then provide the path of length two. Also, $x^*z, y^*z \in E(\mathrm{LAT}(G))$, so $\mathrm{dist}_{\mathrm{LAT}(G)}(x^*, y^*) = 2$.

For a pair of vertices a, b with $a \in \{x^*, y^*\}$ and $b \in \{x, y\}$, whenever a is not the anti-clone of b, we have that if $xy \notin E(G)$ then $ab \in E(\mathrm{LAT}(G))$. Otherwise, we find a new vertex, say w, not adjacent to x nor y. Find a third vertex u such that each of b, w, and u are all pairwise non-adjacent. Again, this must exist since $\gamma(G) \geq 3$. We have a path as shown in Figure 4.6 with vertices $\{a, w, u^*, b\}$ which is a path of length three from a to b. Thus, we find that $\mathrm{dist}_{\mathrm{LAT}(G)}(a, b) \leq 3$.

Therefore, whenever $\gamma(G) \geq 3$, we have that diam $(\mathrm{LAT}(G)) \leq 3$. $\square$

LEMMA 4.11. If G is a graph that is neither K_1 nor $\overline{K_2}$, then neither $\mathrm{LT}(G)$ nor $\mathrm{LAT}(G)$ is a disjoint union of two cliques.

PROOF. If $G = K_2$, then $\mathrm{LT}(G)$ is connected, and $\mathrm{LAT}(G)$ has three components. When $|V(G)| \geq 3$, if either $\mathrm{LT}(G)$ or $\mathrm{LAT}(G)$ is a disjoint union of two cliques then no two clones
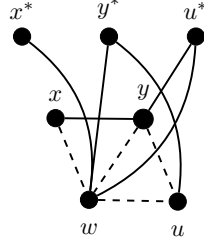
FIGURE 4.6. An illustration of the path of length three in Lemma 4.10.

could be in the same component. Since there are at least three vertices, there are at least three clones, so by pigeonhole, there must be two clones in one component. This is a contradiction. Therefore, neither $LT(G)$ nor $LAT(G)$ is a disjoint union of two cliques. $\square$

We use the previous lemmas for the following theorem on the diameter of the graphs generated by the ILM, and the final theorem In our section of graph parameters.

THEOREM 4.12. *Let $G \neq K_1$ be a graph that is not the disjoint union of two cliques, and $S = (s_0, s_1, s_2, \ldots)$ be a binary sequence with at least two bits equal to 0, where τ is the index of the second time-step such that $s_\tau = 0$, then for all $t \geq \tau + 1$,*

$$\mathrm{diam}\left(ILM_{t,S}(G)\right) = 3.$$

PROOF. Let $G_t = \mathrm{ILM}_{t,S}(G)$ for all $t \geq 0$. Using Lemma 4.3, the radius of G_t for all $t \geq \tau + 1$ is at least three. Therefore, the diameter is at least three, which gives our lower bound.

We next consider the upper bound. Since transitive steps preserve the diameter, we only need to consider the diameter immediately after the most recent anti-transitive step. Without loss of generality assume $s_{t-1} = 0$, thus, $G_t = \mathrm{LAT}(G_{t-1})$. Using Lemma 4.11, we know that G_i is not the disjoint union of two cliques for all $0 \leq i \leq t-1$. It suffices to prove this theorem in the case where $t - 1 = \tau$ is the second instance of a transitive step, and where $s_0 = 0$.

If $\gamma(G_{t-1}) \geq 3$, then we are done by Lemma 4.10. We do not have a dominating vertex at time-step t since we are taking an anti-transitive step, so our remaining case is $\gamma(G_{t-1}) = 2$. Thus, there exist two vertices $u, v \in V(G_{t-1})$ whose closed neighbourhoods partition the vertex set of G_{t-1}. Using Theorem 4.8, G cannot have a dominating vertex since $\gamma(G_{t-1}) = 2$ and there was only one prior instance of an anti-transitive step. Using Lemma 4.11, G_1 is connected. Hence, G_{t-1} must also be connected since we are only taking transitive steps, which preserve connectivity. Similarly, G_{t-1} cannot have a dominating vertex since $\Delta(G_{t-1}) = \frac{n_{t-1}}{2} - 1$. By Lemma 4.11, G_t is connected.

Let x and y be a pair of vertices whose closed neighbourhoods partition the vertex set of G_{t-1}. By Lemma 4.7 they must also partition $V(G_t)$. Thus, every pair of vertices in $N[x]$ are at distance at most two from each other. Consider a vertex $u \in N[x]$. If u has a neighbour in $N[y]$, as shown in Figure 4.7, then for all vertices $v_1 \in N[x]$ we have $\mathrm{dist}(u, v_1) \leq 2$ and for all vertices $v_2 \in N[y]$ we have $\mathrm{dist}(u, v_2) \leq 3$.

We consider pairs of vertices u, v such that $N[u] \subseteq N[x]$ and $N[v] \subseteq N[y]$. There are two cases to consider.

70

Case 1: $u, v \in V(G_{t-1})$

We then have that $\deg(u) = \deg(v) = \frac{n_t}{2} - 1$. There must be some vertex a with a neighbour in each of $N[u]$ and $N[v]$ or else G would be disconnected. Thus, $\mathrm{dist}(u, v) \leq 3$.

Case 2: Without loss of generality, $v \in V(G_t) \setminus V(G_{t-1})$ and $u \in V(G_{t-1})$.

Let $w \in V(G_t) \cap N[x]$ such that the anti-clone of w is $v = w^*$. Since $v = w^*$ is not adjacent to any vertex in $N[x] \cap V(G_{t-1})$, it must be the case that $V(G_{t-1}) \cap N[x] = N[w] \cap V(G_{t-1})$. Hence, $N[w]$ and $N[y]$ partition the vertex set of G_{t-1}, so we may consider $x = w$ and $v = w^* = x^*$, as is described in Figure 4.7. Thus, $N[y] \cap V(G_{t-1}) \subseteq N[x^*]$ since $N[x], N[y]$ partition the vertex set.
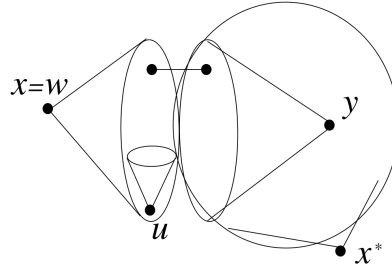
FIGURE 4.7. An illustration of the relationship of vertex partitions in Theorem 4.12.

We know that G_{t-1} is connected, so there must be some edge, say ab, from $N[y] \cap V(G_{t-1})$ to $N[x] \cap V(G_{t-1})$ with endpoint $a \in N[x]$ and endpoint $b \in N[y]$. If $N[u] = N[x]$ we have the path $\{u, a, b, v\}$. Otherwise, $N[u] \subsetneq N[x]$ implies $\deg(u) < \frac{n_t}{2} - 1$ which implies that u is some anti-clone of some vertex of G_{t-1}. We now apply a symmetric argument to u.

Assume, without loss of generality that u is the clone of y. Since u and v are clones of a pair of vertices whose closed neighbourhoods partition the vertex set of G_{t-1} then it must be true that $N[u] = V(G_{t-1}) \backslash N[y] = N[x] \cap V(G_{t-1})$ and also that $N[v] = N[y] \cap V(G_{t-1})$. Since G_{t-1} is connected, we have found a path of length exactly three from u to v. This concludes the proof of the upper bound.

Therefore, whenever G is neither the disjoint union of cliques nor K_1, with a binary sequence that witnesses at least two 0's before time-step τ, then for all $t \geq \tau$, $\text{diam}\left(\text{ILM}_{t,S}(G)\right) = 3$. $\qquad \square$

4.3. Structural Results

In this section, we will focus on two main results. The first result is the Hamiltonicity of the graph, which will define and discuss the existence of Hamilton cycles. We will then provide a result that says every finite graph will eventually appear as an induced subgraph of $\text{ILM}_{t,S}(G)$ as t grows large.

4.3.1. Hamiltonicity. In the subsequent proof, we will use the following edge switch technique. Let G be a graph with two disjoint cycles of arbitrary length, $C_{(1)}$ and $C_{(2)}$, and let C_4 be a 4-cycle that has one edge in common with $C_{(1)}$, say e, and one with $C_{(2)}$, say f. In this case, we call this copy of C_4 an *edge switch* between e and f, and note that an edge switch implies that there is a cycle that covers all the vertices in $V(C_{(1)}) \cup V(C_{(2)})$.

We will also require Dirac's Theorem, which we restate here [23].

THEOREM 4.13. *Every graph G with $|V(G)| = n \geq 3$ and minimum degree $\delta(G) \geq n/2$ has a Hamilton cycle.*

We may now proceed to prove the theorem.

THEOREM 4.14. *Let $G \neq K_1$, and $S = (s_0, s_1, s_2, \ldots)$ a binary sequence with at least two nonconsecutive 0's such that if there are exactly two 0's, there is a 1 in between them. If t_0 is the first index such that $s_{t_0} = 0$ and $s_{t_0-k} = 0$ for some $k \geq 2$, then for all $t \geq t_0 + 1$, $ILM_{t,S}(G)$ is Hamiltonian.*

PROOF. Let $G_t = \text{ILM}_{t,S}(G)$ for any particular $t \geq t_0 + 1$. Let τ_1 be the lowest index such that $s_{\tau_1} = 0$ and let $\tau_1 + 1 \leq \tau_2 \leq t - 1$ be the largest index such that $s_{\tau_2} = 0$. For all $\theta \geq \tau_1 + 1$ we have that $\Delta(G_\theta) = \frac{n_\theta}{2} - 1$ as shown in Theorem 4.8. Thus, $\delta\left(\overline{G_\theta}\right) = \frac{n_\theta}{2}$. Since $|V(G_\theta)| \geq 4$, by Theorem 4.13, we have that $\overline{G_\theta}$ is Hamiltonian.

Let $uv \in E\left(\overline{G_{\tau_2-1}}\right)$ be an edge in the Hamilton cycle, which must exist since τ_1 and τ_2 are nonconsecutive 0's, thus $\tau_1 + 1 < \tau_1 + 2 \leq \tau_2$.

Claim: $\overline{G_{\tau_2}}$ has a Hamilton cycle with four consecutive vertices that form a clique.

The clone vertices of $\overline{G_{\tau_2}}$, the set $V(G_{\tau_2}) \backslash V(G_{\tau_2-1})$, form an independent set of cardinality $n_{\tau_2}/2$, and thus, also form a clique of at least cardinality four in $\overline{G_{\tau_2}}$. Consider the Hamilton cycle in $\overline{G_{\tau_2-1}}$. We traverse a Hamilton path from u to v. The edge $vv^* \in E\left(\overline{G_{\tau_2}}\right)$, where v^* is the anti-clone of v, and we can continue a Hamilton path from v to v^* and then traverse all the anti-clones in $V(G_{\tau_2}) \backslash V(G_{\tau_2-1})$ which form a clique in $\overline{G_{\tau_2}}$, ending with u^*, the anti-clone of u. See Figure 4.8.
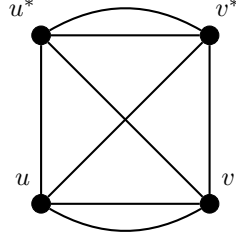
73

FIGURE 4.8. Creating a Hamilton cycle in the complement after one anti-transitive time-step in Theorem 4.14.

In particular, since there are at least four vertices in G_{τ_2-1}, we have at least four anti-clones in a Hamilton cycle in $\overline{G_{\tau_2}}$. Further, these consecutive vertices along a Hamilton cycle form a clique of order four. This completes the proof of the claim.

Let $C = \{v_1, v_2, \ldots, v_{n_{\tau_2}}\}$ be the Hamilton cycle in $\overline{G_{\tau_2}}$. Without loss of generality, let $\{v_1, v_2, v_3, v_4\}$ be the four consecutive vertices that form a clique. Consider now, the clones of each v_i in G_{τ_2+1}, denoted v_i^*, respectively. Since $v_iv_{i+1} \in E\left(\overline{G_{\tau_2}}\right)$, it must be the case that $v_iv_{i+1}^* \in E\left(G_{\tau_2+1}\right)$ and that $v_i^*v_{i+1} \in E\left(G_{\tau_2+1}\right)$ for each $1 \le i \le n_{\tau_2} - 1$ and for the endpoint that $v_1^*v_{n_{\tau_2}}, v_1v_{n_{\tau_2}}^* \in E\left(G_{\tau_2+1}\right)$. These edges form two disjoint cycles in G_{τ_2+1}, one with edges of the form $v_iv_{i+1}^*$, call this $C_{(1)}$ and the other with edges of the form $v_i^*v_{i+1}$, call this $C_{(2)}$, with appropriate modularity for the endpoints. Each cycle has length $n_{\tau_2+1}/2$, is disjoint and $V(C_{(1)}) \cup V(C_{(2)}) = V(G_{\tau_2+1})$.

Since $\{v_1, v_2, v_3, v_4\}$ form an independent set in G_{τ_2}, we know that $v_1v_3^*, v_2^*v_4 \in E(G_{\tau_2+1})$. Using the edge switch technique as defined earlier. We have that $\{v_1v_2^*, v_2^*v_4, v_4v_3^*, v_3^*v_1\}$ is a four cycle, where exactly one edge is in $C_{(1)}$ and exactly one distinct edge is in $C_{(2)}$. We

74

may then take an edge switch to form a Hamilton cycle on the vertices $V(G_{\tau_2+1})$ as depicted in Figure 4.9. Hence, G_t is Hamiltonian, whenever the sequence S witnesses at least two non-consecutive 0's. $\square$
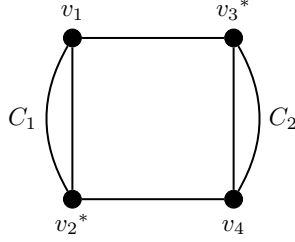

FIGURE 4.9. Using the edge switch technique to create a Hamilton cycle in Theorem 4.14.

For a sequence of all 0's, we have the property that there are at least two nonconsecutive 0's. Thus, we have the following corollary.

COROLLARY 4.15. *For any graph G when $t \geq 3$, then $ILAT_t(G)$ is Hamiltonian.*

Consider next the Iterated Local Transitivity model. Since we never experience anti-transitive steps, we cannot utilize the previous theorem or proof techniques. In fact, we notice that it may take large values of t before $\mathrm{ILT}_t(G)$ becomes Hamiltonian for some graph G. We now introduce the following parameter.

DEFINITION 4.1. Let G be a connected graph. We define $\zeta(G)$ to be the smallest integer such that for all $t \geq \zeta(G)$ that $\mathrm{ILT}_t(G)$ is Hamiltonian. This is proved to be well-defined in Theorem 4.16. We define

75

$$\zeta_n = \max\{\zeta(G) : |V(G)| = n, G \text{ is connected}\}.$$

The next theorem shows that ζ_n grows with n.

THEOREM 4.16. *For all $n \geq 3$, we have that*

$$\log_2(n-1) \leq \zeta_n \leq \lceil \log_2(n-1) \rceil + 1.$$

PROOF. We begin with the lower bound. Consider the star $K_{1,n-1}$ on n vertices. Let A_t denote the set of all vertices in $\text{ILT}_t(K_{1,n-1})$ that are descendants of the centre vertex; that is they are clones of the centre, x, at each time-step, including the clones of clones. It is the case that A_t is a vertex-cut of cardinality 2^t whose removal leaves a graph with at least $n-1$ components. If $2^t < n-1$, then $\text{ILT}_t(K_{1,n-1})$ cannot be Hamiltonian. Therefore, $t \geq \log_2(n-1)$ is a lower bound for ζ_n.

We continue with the upper bound. Fix a graph G on n vertices and let $G_t = \text{ILT}_t(G)$. If G_t is Hamiltonian, then let $C = \{v_1, v_2, \ldots, v_{n_t}\}$ be the Hamilton cycle in G_t. In G_{t+1}, we can then find a Hamilton cycle by replacing the edges of the form $v_i v_{i+1}$ with the edges $v_i v_i'$ and $v_i' v_{i+1}$ for $1 \leq i \leq n_t - 1$, and replacing the edge $v_{n_t} v_1$ with the edges $v_{n_t} v_{n_t}'$ and $v_{n_t}' v_1$ for the endpoint. Hence, transitive steps preserve the Hamiltonicity of the graph.

It suffices to now find the first t such that G_t is Hamiltonian since it will continue to be Hamiltonian for all subsequent time-steps.

Claim 1: For all $t \geq 2$, $\mathrm{ILT}_t(K_1)$ contains a Hamilton cycle in which each vertex of $\mathrm{ILT}_{t-1}(K_1)$ is adjacent to its clone in $\mathrm{ILT}_t(K_1)$.

To prove this claim we will use induction. Our base case considers $\mathrm{ILT}_2(K_1)$, which has a Hamilton cycle that can be found by inspection in Figure 4.10. The inductive step

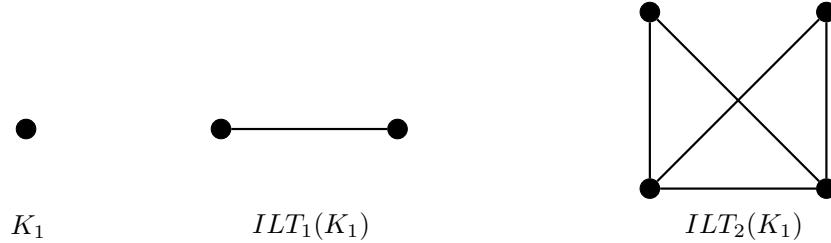K_1 $ILT_1(K_1)$ $ILT_2(K_1)$

FIGURE 4.10. The first two time-steps of the ILT model on K_1 demonstrating the existence of the Hamilton cycle in Theorem 4.16.

is the argument as described in the above paragraph. Once a Hamilton cycle is found, we can continue to build a Hamilton cycle in which each clone is adjacent to its parent. This concludes the proof of the claim.

Claim 2: If $V(K_1) = \{v\}$ and $V(\mathrm{ILT}_1(K_1)) = \{u, v\}$, then for all $t \geq 1$, $\mathrm{ILT}_t(K_1)$ contains a perfect matching that matches the descendants of u to the descendants of v that are not descendants of u.

For $t \geq 2$, one such matching has the property that the edges can be paired off in such a way that if e and f are paired, there exist two vertices, $x, y \in V(\mathrm{ILT}_{t-1}(K_1))$ such that $e = xy'$ and $f = x'y$. We refer to these edges as a *paired matching*, and indicate the relationship in Figure 4.11.
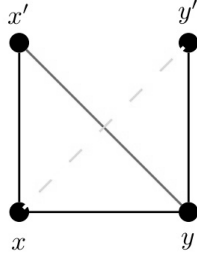
FIGURE 4.11. An example of a paired matching in Theorem 4.16.

We proceed by induction. The base case holds as shown previously. Assume true for some $\tau > 1$. Let $M_{\tau-1}$ be the desired matching in $\text{ILT}_{\tau-1}(K_1)$. Let M_τ then be defined as

$$M_\tau = \{xy' : xy \in M_{\tau-1}\}.$$

This concludes the proof of Claim 2.

Let T be a spanning tree of G, and let $t_0 = \lceil \log_2(\Delta(T)) \rceil + 1$. We will show that G_{t_0} is Hamiltonian. Let $V(T) = \{v_1, v_2, \ldots, v_{n_0}\}$, and let

$$V_{i,t} = \{v \in V(G_t) : v \text{ is a descendant of } v_i\}.$$

Observe that for all $t \geq 0$, $G_t[V_{i,t}] = \text{ILT}_t(K_1)$ and if $v_i v_j \in E(T)$, then $G_t[V_{i,t} \cup V_{j,t}] = \text{ILT}_{t+1}(K_1)$. By Claim 1, for each $1 \leq i \leq n_0$, we can find a cycle $C_{(i)}$ that covers $V_{i,t}$, and has the property that the vertices in $V_{i,t-1}$ are adjacent to their clones in the cycle.

By Claim 2, for each edge $v_i v_j \in E(T)$, we can find a paired matching from V_{i,t_0} to V_{j,t_0}. For each edge $v_1 v_j \in E(T)$, given any vertex $v \in V_{i,t_0-1}$, there exists some $u \in V_{j,t_0-1}$ such

78

that there is an edge switch between vv' and uu'. This property guarantees that we can build a Hamilton cycle by iteratively using edge switches between edges of the form uu' and vv', one for each edge in T. As long as we never perform two edge switches at the same vertex, this will give us a Hamilton cycle, and we now describe that it would not be possible to perform two edge switches at the same vertex. Since $t_0 \geq \log_2(\Delta(T)) + 1$, we have that $|V_{i,t_0-1}| = 2^{t_0-1} \geq \Delta(T)$. Thus, we can avoid performing two edge switches at the same vertex, so G_{t_0} is Hamiltonian. Since $n_0 - 1 \geq \Delta(T)$, we have that $\zeta_n \leq \lceil \log_2(n-1) \rceil + 1$ and this completes the proof. □

4.3.2. Induced Subgraphs. We conclude the chapter with a result on induced subgraphs. For any finite graph F, given any graph G and any binary sequence S we can find some $\tau \geq 0$ such that for all $t \geq \tau$, F appears as an induced subgraph of $\text{ILM}_{t,S}(G)$.

THEOREM 4.17. *If F is a graph, then there exists some constant $t_0 = t_0(F)$ such that for all $t \geq t_0$, all graphs G, and all binary sequences S, F is an induced subgraph of $\text{ILM}_{t,S}(G)$.*

PROOF. To begin, we will first show for some constant $k = k(F)$ that $\text{ILT}_k(K_1)$ contains an induced copy of F. We will then show for large enough constant $t_0 = t_0(k)$ that $\text{ILM}_{t,S}(G)$ contains an induced copy of $\text{ILM}_k(K_1)$ for all binary sequences S and graphs G. Let $|V(F)| = \ell$.

Claim: $\text{ILT}_{\ell + \binom{\ell}{2} - 1}(K_1)$ contains an induced copy of F.

We remark that the clique number, $\omega(\text{ILT}_{\ell-1}(K_1)) \geq \ell$, since at each time-step we increase the order of the largest clique by at least one. To show this is true, consider any

79

maximum order clique and one vertex of this clique. The clone of this particular vertex will be adjacent to all vertices in the clique, and so increases the order of the largest clique by at least one.

We now proceed to show that, for an arbitrary graph H if H is an induced subgraph of $\mathrm{ILT}_r(K_1)$, then $H - e$ is an induced subgraph of $\mathrm{ILT}_{r+1}(K_1)$ for any edge $e \in E(H)$. Let $e = uv \in E(H)$ and let $V_r(H)$ be the set of vertices in $\mathrm{ILT}_r(K_1)$ that induce a copy of H. Let u', v' be the clones of the vertices u, v, respectively. We then have that $(V(H) \backslash \{u, v\}) \cup \{u', v'\}$ induces a copy of $-e$ since $N_r[u] = N_{r+1}(u')$ and $N_r[v] = N_{r+1}(v')$, and also that u', v' are not adjacent, as can be seen in Figure 4.12.
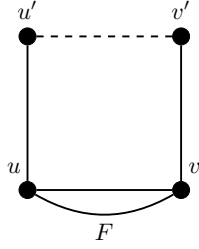
FIGURE 4.12. The removal of edge $e = uv$ from the induced subgraph of H in Theorem 4.17.

For any graph F on ℓ vertices we can find a time-step $\tau \geq \ell - 1$ such that $\mathrm{ILT}_\tau(K_1)$ contains a copy of K_ℓ, the complete graph on ℓ vertices. We then inductively remove edges from this copy of K_ℓ until we have found an induced copy of F. We need one time-step for each edge removal, so we need at most $\binom{\ell}{2}$ time-steps to find this induced subgraph.

Therefore, we have shown that $\mathrm{ILT}_{\ell-1+\binom{\ell}{2}}(K_1)$ contains an induced copy of every graph on ℓ vertices.

We continue by showing that for any fixed k, for any graph G, and for any binary sequence S that $\mathrm{ILM}_{2k,S}(G)$ contains an induced copy of $\mathrm{ILT}_k(K_1)$. If $\mathrm{ILM}_{t,S}(G)$ for some $t \geq 0$ contains a copy of $\mathrm{ILT}_r(K_1)$ for any integer r when a transitive step is performed then $\mathrm{ILM}_{t+1,S}(G)$ contains an induced copy of $\mathrm{ILT}_{r+1}(K_1)$, by considering the vertices that induced the copy of $\mathrm{ILT}_r(K_1)$ along with their clones in time-step $t+1$.

We claim that if $\mathrm{ILM}_{t,S}(G)$ contains an induced copy of $\mathrm{ILT}_r(K_1)$ and we perform two anti-transitive steps, then $\mathrm{ILM}_{t+2,S}(G)$ contains an induced copy of $\mathrm{ILT}_{r+1}(K_1)$.

For every vertex $x \in V(\mathrm{ILM}_{t,S}(G))$, let $x^{**} \in V(\mathrm{ILM}_{t+2,S}(G)) \setminus V(\mathrm{ILM}_{t+1,S}(G))$ denote the anti-clone of the anti-clone of x. We use $N_t[x]$ to denote the closed neighbourset of x at time step t for any $t \geq 0$. We have the following:

$$N_{t+2}[x^{**}] = V(\mathrm{ILM}_{t+1,S}(G)) \setminus N_{t+1}[x^*]$$
$$= V(\mathrm{ILM}_{t+1,S}(G)) \setminus (V(\mathrm{ILM}_{t,S}(G)) \setminus N_t[x])$$
$$= ((V(\mathrm{ILM}_{t+1,S}(G)) \setminus V(\mathrm{ILM}_{t,S}(G)))) \cup N_t[x].$$

Thus, $N_t[x] \subseteq N_{t+2}[x^{**}]$. In fact, $N_t[x] = N_{t+2}[x^{**}] \cap V(\mathrm{ILM}_{t,S}(G))$. In fact, $N_t[x] = N_{t+2}[x^{**}] \cap V(\mathrm{ILM}_{t,S}(G))$. So, if $X \subseteq V(\mathrm{ILM}_{t,S}(G))$ induces a copy of $\mathrm{ILM}_r(K_1)$, then $X \cup (X^{**} \cap V(\mathrm{ILM}_{t,S}(G)))$ induces a copy of $\mathrm{ILM}_{r+1}(K_1)$. Thus, for any clique of largest order in $\mathrm{ILM}_{t,S}(G)$ that contains x as a vertex of the clique, then together with x^{**} we have that $\mathrm{ILM}_{t+2,S}(G)$ contains a clique of order one larger. Therefore, whenever $\mathrm{ILM}_{t,S}(G)$

contains a copy of $\text{ILM}_r(K_1)$, we have that $\text{ILM}_{t+2,S}(G)$ contains a copy of $\text{ILM}_{r+1}(K_1)$ if $s_{t+1} = s_{t+2} = 0$.

Thus, any combination of k transitive steps or pairs of subsequent anti-transitive steps will yield and induced copy of $\text{ILM}_k(K_1)$. Observe that such a sequence must occur within the first $2k$ time-steps. If we partition the first $2k$ elements of the sequence S into k contiguous pairs, then each pair either contains a one or is a pair of zeros, which will yield the appropriate structure as described above. This completes the proof. $\square$

CHAPTER 5

Iterated Global Models

5.1. Introduction

The previous models discussed in the thesis including ILT, ILAT, and ILM focused on considering the local structure of the graph and generating a new model iteratively from this structure. In this chapter, we now define a model that is independent of the structure of the initial graph but retains the iterative character of the previously defined models. We introduce the *Iterated Global Models*, where a dominating vertex is added for each subset of vertices of a given cardinality. This creates large graphs that behave similarly to unbalanced complete bipartite graphs, which will be discussed in subsequent sections.

Let $k \geq 1$ be an integer. The one parameter of the model is the initial, connected graph $G = G_0$. At each time-step $t \geq 0$, we create G_{t+1} from G_t in the following way: for each set of vertices of cardinality $\lfloor \frac{1}{k} n_t \rfloor$, say S, add a new v_S that is adjacent to each vertex of S. We name this process the $\frac{1}{k}$-*model*. For ease of notation and for consistency with earlier chapters, we refer to newly added vertices in G_{t+1} as *clones*. Note that the clones form an independent set in G_{t+1}.

We will focus in this chapter on the case $k = 2$, which we refer to as the *half-model*. In the half-model, each new vertex is adjacent to approximately half of the existing network. See Figure 5.1 for an example.
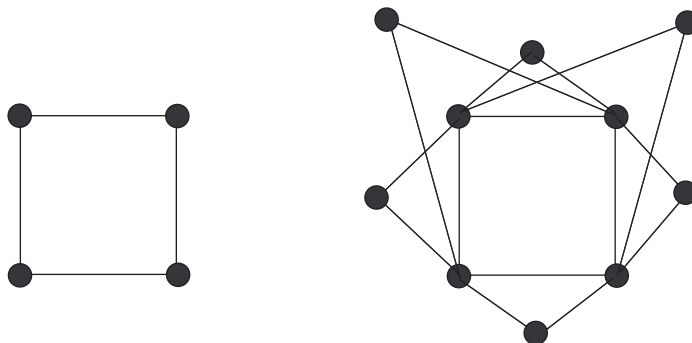
FIGURE 5.1. One time-step of the half-model beginning with C_4.

One motivation for the model is that a new vertex chooses a prescribed fraction of the new vertices to link to, without preference of local or other structures. For example, a user on social media may choose to like a fraction of the posts on their newsfeed. The half-model takes this assumption to an extreme to provide structural insight into this deterministic process.

5.2. Complex network properties

Our first result establishes the order and size of graphs generated by the half-model. We first recall Stirling's approximation for the factorial:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

Stirling's approximation may be used to derive an expression for the central binomial coefficient,

$$\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}}.$$

which may be derived directly and is part of folklore. This approximation will be useful in our analysis, and its usefulness has provided motivation for the study of the half-model as opposed to other values of k.

For an exposition of the asymptotics of binomial coefficients, see the book *Asymptopia* by Spencer and Florescu [22].

THEOREM 5.1. *The order and size of the graph G_t in the half-model are given by the following, respectively:*

$$n_t \sim \binom{n_{t-1}}{\lfloor \frac{n_{t-1}}{2} \rfloor} \qquad and \qquad e_t \sim \binom{n_{t-1}}{\lfloor \frac{n_{t-1}}{2} \rfloor} \cdot \lfloor \frac{n_{t-1}}{2} \rfloor.$$

To simplify notation as we proceed with the chapter, we define the function:

$$\alpha_t = \binom{n_t}{\lfloor \frac{n_t}{2} \rfloor}.$$

PROOF. We begin with the order of G_t. By the definition of the model, at each time-step $t \geq 1$, we add one vertex for each set of size $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$. Hence, we derive the following sum:

$$n_t = n_0 + \sum_{i=1}^{t} \alpha_{i-1}.$$

The term α_{t-1} will dominate the rest of the summation, which gives us the desired expression for the order of G_t.

Next, we determine the size of G_t. Each new vertex added is adjacent to a set of size $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$, and we add α_{t-1} vertices, so we obtain the following recursive formula for the number of edges at time-step t :

$$e_t = e_{t-1} + \left\lfloor \frac{n_{t-1}}{2} \right\rfloor \alpha_{t-1}.$$

We observe that the second term dominates the sum, and the result follows. $\square$

The half-model gives rise to graphs with average degree tending to infinity.

THEOREM 5.2. *The half-model densifies with time.*

PROOF. By Theorem 5.1, we have that

$$\frac{e_t}{n_t} \sim \frac{\alpha_{t-1} \cdot \left\lfloor \frac{n_{t-1}}{2} \right\rfloor}{\alpha_{t-1}} = \left\lfloor \frac{n_{t-1}}{2} \right\rfloor,$$

which tends to infinity with t. $\square$

The half-model has bad spectral expansion, as in the case of the ILM model. As in Chapter 3, we use the following lemma.

LEMMA 5.3 ([**12**]). *If G is a graph with spectral gap λ, then for all sets $X \subseteq V(G)$,*

$$\left| e(X, X) - \frac{(\text{Vol}(X))^2}{\text{Vol}(G)} \right| \leq \lambda \frac{\text{Vol}(X)\text{Vol}(\overline{X})}{\text{Vol}(G)}.$$

THEOREM 5.4. *Graphs generated by the half-model satisfy $\lambda_t \sim 1$, where λ_t is the spectral gap of G_t.*

PROOF. Let $X = V(G_t) \backslash V(G_{t-1})$ be the set of cloned vertices added to G_{t-1} to form G_t. Since X is an independent set, we note that $e(X, X) = 0$. We derive that

$$\text{Vol}(G_t) = 2e_t \sim \alpha_{t-1} \cdot n_{t-1},$$

$$\text{Vol}(X) = \alpha_{t-1} \cdot \left\lfloor \frac{n_{t-1}}{2} \right\rfloor,$$

$$\text{Vol}(\overline{X}) \sim \alpha_{t-1} \cdot \left\lfloor \frac{n_{t-1}}{2} \right\rfloor.$$

87

Hence, by Lemma 5.3, we have that

$$
\begin{aligned}
\lambda_t &\geq \frac{(\mathrm{Vol}(X))^2}{\mathrm{Vol}(G_t)} \cdot \frac{\mathrm{Vol}(G_t)}{\mathrm{Vol}(X)\mathrm{Vol}(\overline{X})} \\
&= \frac{\mathrm{Vol}(X)}{\mathrm{Vol}(\overline{X})} \\
&\sim \frac{\alpha_{t-1} \cdot \left\lfloor \frac{n_{t-1}}{2} \right\rfloor}{\alpha_{t-1} \cdot \left\lfloor \frac{n_{t-1}}{2} \right\rfloor} \\
&= 1,
\end{aligned}
$$

and the result follows. $\square$

5.2.1. Diameter. We observe that the half-model has a small (in fact, constant) diameter as required for the small-world property. We first prove some results about the connectivity for graphs generated by this model.

LEMMA 5.5. For all $t \geq 0$, if G_t is connected and $n_t \geq 2$, then G_{t+1} is connected.

PROOF. If v is a clone in G_{t+1}, then since $n_t \geq 2$, we have that v is adjacent to at least one vertex u in $V(G_t)\backslash V(G_{t+1})$. Since G_t is connected by hypothesis, there exists a path from u to any other vertex of G_t, and hence, there is such a path from v to any vertex of G_t. Since the vertex v was an arbitrary clone, we have shown there exists a path between any two vertices in G_{t+1}. $\square$

In the case where $n_0 = 1$, then G_0 is K_1. Note that G_1 is $\overline{K_2}$, and G_2 is the disjoint union of two edges. In particular, G_1 and G_2 are not connected. The subsequent lemma

88

will provide insight into how many iterations a disconnected graph requires before becoming connected.

LEMMA 5.6. For all $t \geq 0$, if G_t is a graph with $n_t \geq 4$, then G_{t+1} is connected.

PROOF. We proceed by a proof by contraposition. Suppose then that G_{t+1} is disconnected, and so there exists two vertices u, v in G_{t+1} such that there is no path between them.

Case 1: u, v are both in $V(G_t)$.

In this case, there is no set of size $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$ that contains both u and v, since otherwise, a clone in G_{t+1} would be adjacent to both u, v. At each time-step t, we add a clone for every subset of size $\left\lfloor \frac{n_t}{2} \right\rfloor$; hence, it must be the case that $\left\lfloor \frac{n_t}{2} \right\rfloor < 2$ which implies $n_t \leq 3$. This satisfies the negation of the predicate, and we have proved the result in this case.

Case 2: Exactly one of u or v is not in $V(G_t)$; without loss of generality, say $u \in V(G_{t+1}) \backslash V(G_t)$.

As u is a clone it has degree $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$, and so has a neighbour x in G_t, whenever $n_t \geq 2$. Thus, there is no path from x to v in G_t, and we apply Case 1 using these two vertices.

Case 3: Both u, v are in $V(G_{t+1}) \backslash V(G_t)$.

Since there are at least two clones it must be the case that $\alpha_t \geq 2$, and so $n_t \geq 2$. There then exists some neighbour x of u in G_t and some neighbour y of v in G_t. We then have that there is no path from x to y in G_t and we apply Case 1 to these two vertices. The proof follows. $\square$

Our next result proves the 2-connectivity of graphs generated by the half-model.

LEMMA 5.7. The graph G_t is 2-connected whenever $t \geq 4$, regardless of the input graph G_0.

PROOF. Using the recursive formula for the number of edges at time t in the proof of Theorem 5.1, for any graph G_0, we have at least four vertices after two time-steps. Using Lemma 5.6, we require at least one additional time-step to ensure connectivity. Thus, regardless of the input graph G_0, it is the case that G_t is connected for $t \geq 3$. We now claim that whenever a graph G_t is connected, G_{t+1} will be 2-connected.

Claim: If G_t is connected and $n_t \geq 4$, then G_{t+1} is 2-connected.

If G_t is 2-connected, then we are done since every vertex in $V(G_{t+1})\backslash V(G_t)$ has at least one neighbour in $V(G_t)$, and we may use the same two paths between those neighbours to find 2-connectivity. Suppose G_t is at most 1-connected and thus let u be a cut-vertex of G_t. Consider two vertices in G_t, say a, b, that have a shortest path through u. In G_{t+1}, there is some clone z that is adjacent to both a, b. Therefore, we have two paths from a to b, and the proofs of the claim and theorem follow. $\square$

Our main result of this section is the following.

THEOREM 5.8. *Suppose that G_0 has order at least 4. In the half-model, the diameter of G_t for $t \geq 5$, is at most three.*

90

PROOF. We consider the distance between two non-adjacent vertices $x, y \in V(G_t)$ in three cases.

Case 1: $x, y \in V(G_{t-1})$.

There exists some set $S \subseteq V(G_{t-1})$ of cardinality $\lfloor \frac{n_{t-1}}{2} \rfloor$ containing both x and y. Thus, the dominating vertex for this set S, v_S is adjacent to both x and y so their distance is 2.

Case 2: $x \in V(G_{t-1})$ and $y \notin V(G_{t-1})$.

There exists a vertex $z \in N_{G_t}(y)$. There is some set $S \subseteq V(G_{t-1})$ so that $x, z \in S$. The vertex v_S that dominates S in G_t is adjacent to both x and z, so we have the path $y z v_S x$. Hence, the distance between x and y is at most 3. The symmetric case where $y \in V(G_{t-1})$ and $x \notin V(G_{t-1})$ is analogous.

Case 3: $x, y \notin V(G_{t-1})$.

Since x, y are new vertices in time-step t, there must be two sets $S_x, S_y \subseteq V(G_{t-1})$, where x dominates S_x and y dominates S_y. If $S_x \bigcap S_y \neq \emptyset$, then there is some vertex of G_{t-1} adjacent to both x and y, so their distance is 2. Suppose now that $S_x \bigcap S_y = \emptyset$. Since $|S_x| = |S_y| = \lfloor \frac{n_{t-1}}{2} \rfloor$, it may be the case that there exists a vertex $z \notin S_x \cup S_y$.

Suppose first that there is no such vertex z. There must be some edge with one endpoint in S_x and the other in S_y, since otherwise, the graph would be disconnected, which contradicts Lemma 5.7. We call these two endpoints a and b. We then have a path $xaby$ and the distance between x and y is 3.

91

If there is such a vertex z, then since G_t is 2-connected by Lemma 5.7, z cannot be a cut-vertex. Therefore, there must be some edge with one endpoint in S_x and the other in S_y and the distance between x and y is 3. $\square$

5.3. Graph Parameters for the Half-Model

In this section, we discuss classical graph parameters for the half-model. We begin by considering the independence and clique number.

THEOREM 5.9. *The independence number of G_t is α_{t-1} and for the clique number we have*

$$\chi(G_t) \geq \min\left(\left\lfloor \frac{n_{t-1}}{2} \right\rfloor + 1, \omega(G_0) + t\right).$$

PROOF. At each time-step t, all the cloned vertices form an independent set. The set of new vertices has order $\alpha_{t-1} \geq n_{t-1}$, so this set must be the largest independent set in G_t. Therefore, we derive that $\alpha(G_t) = \alpha_{t-1}$.

We next consider the clique number of G_t. At each time-step t, we add a dominating vertex to subsets of cardinality $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$ from G_{t-1}. If the largest clique K at time-step $t-1$ is contained in one such subset, then we have increased the order of K by 1. However, the maximum degree of new vertices is $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$. Hence, we cannot increase the size of the largest clique to be larger than $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor + 1$. $\square$

We next give the chromatic number of the half-model.

THEOREM 5.10. *For the half-model, we have that the chromatic number is given by*

$$\chi(G_t) = \min\left(\chi(G_0) + t, \left\lfloor \frac{n_{t-1}}{2} \right\rfloor + 1\right).$$

PROOF. Suppose that G_t is properly coloured. Consider a *rainbow* subset of vertices; that is, a set of vertices that requires each distinct colour in the graph. Let the cardinality of this set be $r \geq 1$. When $r \leq \left\lfloor \frac{n_{t-1}}{2} \right\rfloor$, any new clone added that contains this set in its neighbours will need a new colour. When $r > \left\lfloor \frac{n_{t-1}}{2} \right\rfloor$, any new clone that is added will have a neighbour set smaller than the cardinality of the colours, which implies there will always be an available colour. □

We finish by proving a result on the domination number of graphs generated by the half-model.

THEOREM 5.11. *The domination number of G_t is*

$$\gamma(G_t) = \left\lceil \frac{n_{t-1}}{2} \right\rceil + 1.$$

PROOF. We will first establish the upper bound

$$\gamma(G_t) \leq \left\lceil \frac{n_{t-1}}{2} \right\rceil + 1.$$

Consider a set S of $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$ non-clone vertices in G_{t-1}. The vertex x_S dominates S. The complement T of S in $V(G_{t-1})$ has cardinality $\left\lceil \frac{n_{t-1}}{2} \right\rceil$. Hence, $\{x_S\} \cup T$, is the desired dominating set.

93

For the lower bound, we must show that $\gamma(G_t) > \left\lceil \frac{n_{t-1}}{2} \right\rceil$. For a contradiction, suppose that some set of $\left\lceil \frac{n_{t-1}}{2} \right\rceil$-many vertices, say X, dominates G_t. Suppose first that X consists of non-clones. Regardless of the choice of X, there will be some set of non-clones, call it T, of size $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$ such that $X \cap T = \emptyset$. Thus, x_T is not dominated, which is a contradiction.
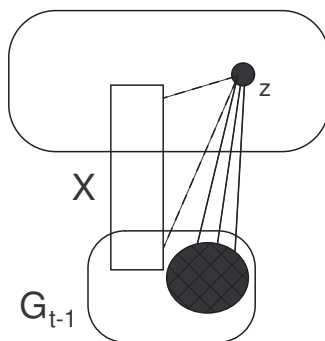

FIGURE 5.2. The vertex z is not adjacent to X.

Suppose that X contains at least one clone. There is a least one clone z not adjacent to $X \cap V(G_{t-1})$, since $|X \cap V(G_{t-1})| < \left\lceil \frac{n_{t-1}}{2} \right\rceil$. See Figure 5.2. Any clone in X is not adjacent to z, since the clones form an independent set. Therefore, z is not dominated by X, which gives a contradiction. $\square$

CHAPTER 6

Conclusion

6.1. Introduction

In this thesis, we introduced two new deterministic models for social networks. The first chapter motivated the models and gave requisite background on graph theory and complex networks. In Chapter 2, we provided background on models that inspired our work: the Iterated Local Transitivity model and the Iterated Local Anti-Transitivity model. We discussed the results for these models, which motivated the new models.

We defined the Iterated Local Model (ILM) in Chapter 3, where we discovered results pertaining to its complex network properties. Specifically, we showed that ILM graphs exhibit the small-world property and densification. Further, the spectral gap was discussed and it was shown that the model has bad spectral expansion as found in real-world social networks.

Chapter 4 focused on the structural properties of ILM graphs, including distance properties and Hamiltonicity. There are many interesting structural properties of this model; in particular, we proved that for any finite graph F, there is a time-step $t \geq 0$ such that the ILM graphs at time-step greater than t will contain the graph F as an induced subgraph.

In Chapter 5, we defined the Iterated Global Model (IGM). The IGM is independent of the original graph structure, unlike ILM graphs, and relies on subsets of vertices to generate new graphs. We proved that graphs generated by the half-model exhibit densification and bad spectral expansion, and we investigated various graph parameters for this model.

6.2. Open problems and further directions

Several open problems remain concerning ILM and half-model graphs. We conclude the thesis by presenting a list of them.

(1) Much less is known about the eigenvalues of the ILAT model than for ILT graphs, and hence, for ILM graphs. In addition, we do not have a precise asymptotic value for the clustering coefficient of ILM graphs.

(2) Rather than cloning vertices, we could instead clone entire families of subgraphs such as paths or triangles. It would be interesting to analyze complex network properties such as the clustering coefficient in this model, as well as classic graph theory parameters.

(3) Randomization of the models would make them more realistic but more complicated to analyze. For instance, in the half-model, we may clone sets of cardinality $\left\lfloor \frac{n_{t-1}}{2} \right\rfloor$ chosen uniformly at random from all possible subsets of G_{t-1}. Duplication models [13], where we randomize which edges to clone, are difficult to rigorously analyze given their dependency structure. The ILM could also be randomized by considering random binary sequences.

96

(4) Graph limits consider dense sequences of graphs and analyze their properties based on their homomorphism densities. Since the ILM and half-models generate dense sequences of graphs, it would be interesting to explore their graph limits.

(5) Many questions remain in the half-model. For example, it would be interesting to estimate the clustering coefficient. Another interesting direction would be to generalize our results to integers $k > 2$.

Bibliography

[1] W. Aiello, F. Chung, L. Lu, A random graph model for massive graphs, *Experimental Mathematics* **10** (2001) 53–66.

[2] N. Alon, J. Spencer, *The Probabilistic Method*, Wiley, New York, 2000.

[3] A. Barabási, R. Albert, Emergence of scaling in random networks, *Science* **286** (1999) 509–512.

[4] V. Boginski, S. Butenko, P.M. Pardalos, On structural properties of the market graph, In: A. Nagurney, editor, *Innovation in Financial and Economic Networks*, Edward Elgar Publishers, pp. 29–45.

[5] B. Bollobás, O. Riordan, The diameter of a scale-free random graph, *Combinatorica* **24** (2004) 5–34.

[6] B. Bollobás, O. Riordan, J. Spencer, G. Tusnády, The degree sequence of a scale-free random graph process, *Random Structures and Algorithms* **18** (2001) 279-290.

[7] A. Bonato, *A Course on the Web Graph*, American Mathematical Society, Providence, Rhode Island, 2008.

[8] A. Bonato, H. Chuangpishit, S. English, B. Kay, E. Meger, The iterated local model for social networks, arXiv:1903.04523 (2019).

[9] A. Bonato, N. Hadi, P. Horn, P. Prałat, C. Wang, Models of on-line social networks, *Internet Mathematics* **6** (2011) 285–313.

[10] A. Bonato, E. Infeld, H. Pokhrel, P. Prałat, Common adversaries form alliances: modelling complex networks via anti-transitivity, In: *Proceedings of WAW'17*, 2017.

[11] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, Graph structure in the web, *Computer Networks* **33** (2000) 309-320.

[12] F.R.K. Chung, *Spectral Graph Theory,* American Mathematical Society, Providence, Rhode Island, 1997.

[13] F.R.K. Chung, L. Lu, *Complex Graphs and Networks*, American Mathematical Society, Providence, Rhode Island, 2004.

[14] D. Easley, J. Kleinberg, *Networks, Crowds, and Markets Reasoning about a Highly Connected World*, Cambridge University Press, 2010.

[15] P. Erdős, A. Rényi, On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5** (1960) 17–61.

[16] E. Estrada, Spectral scaling and good expansion properties in complex networks, *Europhys. Lett.* **73** (2006) 649–655.

[17] J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: densification Laws, shrinking diameters and possible explanations, In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2005.

[18] L. Lovász, *Large networks and graph limits*, American Mathematical Society, Providence, RI, 2012.

[19] V. Memišević, T. Milenković, N. Pržulj, An integrative approach to modeling biological networks, *Journal of Integrative Bioinformatics*, **7**:120, 2010.

[20] J.P. Scott, *Social Network Analysis: A Handbook*, Sage Publications Ltd, London, 2000.

[21] L. Small, O. Mason, Information diffusion on the iterated local transitivity model of online social networks, *Discrete Applied Mathematics* **161** (2013) 1338--1344.

[22] J. Spencer, L. Florescu, Asymptopia, American Mathematical Society, Providence, Rhode Island, 2014.

[23] D.B. West, *Introduction to Graph Theory, 2nd edition*, Prentice Hall, 2001.

[24] W.W. Zachary, An information flow model for conflict and fission in small groups, *Journal of Anthropological Research* **33** (1977) 452–473.