Modelling Negative Correlation in Complex Networks Via Anti-Transitivity

by

Hari Prasad Pokhrel

A thesis presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Science

in the program of

Applied Mathematics

Toronto, Ontario, Canada, 2017

© Hari Prasad Pokhrel, 2017

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

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

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

I further authorize Ryerson University to reproduce this thesis 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 thesis may be made electronically available to the public.

Acknowledgements

I dedicate this work to my beloved father Mr. Rewati Pokhrel. Although he is no longer with us, he always believed in my ability to be successful in academia. You are gone but you left fingerprints of grace in my life.

I would like to thank my mother Ms. Radhika Pokhrel and my wife Ms. Bandana Pokhrel first and foremost for their continued support throughout my mathematics education, and for encouraging me to believe in the power of my dreams.

I am thankful for the support from my thesis supervisor Dr. Anthony Bonato, throughout my master's degree education. I will be graduating this academic year thanks in part to his excellent mentoring and guidance in conducting and finishing my thesis.

I appreciate to Dr. Dejan Delić and Dr.Konstantinos Georgiou for their time and energy in participating in my thesis committee. I also thank Dr. Ewa Infeld for her support with aspects of the thesis. I acknowledge NSERC for partially funding this thesis.

Abstract

Modeling Negative Correlation in Complex Networks Via Anti-Transitivity

Hari Prasad Pokhrel

MSc Applied Mathematics

2017

Ryerson University

Negative correlation appears often in complex networks. For example, in social networks, negative correlation corresponds to rivalry between agents in the network, while in stock market graphs, negative correlation corresponds stocks that move in opposite directions in price action.

We present a simplified, deterministic model of negative correlation in networks based on the principle of anti-transitivity: a non-friend of a non-friend is a friend. In the Iterated Local Anti-Transitivity (ILAT) model, for every node u in a given time-step, we add an anti-clone node that is adjacent to the complement of the closed neighborhood of u. We prove that graphs generated by the ILAT model satisfy several properties observed in complex networks, such as high density and densification power laws, constant diameter, and high local clustering. We also prove that the domination and cop numbers of graphs generated by the ILAT model are bounded above by small, absolute constants as time increases.

Contents

List of Figures

CHAPTER 1

Introduction

1.1. Background and Motivation

Formal properties of social networks were first studied by Simmel and others in the 1930s; see [**6**]. They adopted the terminology of graphs to describe social relations. In the 1950s, American researchers led by Harrison White developed a formal methodology for social network analysis; see [**6**] for additional background. Social network research has grown in intensity to modern times; see [**8, 10**] who established measures for social networks from the 1950s to 1970s.

With the invention and adoption of the internet, and the growth of social, biological, financial and other networks, the use of networks has become even more popular and been applied to a variety of fields. Many researchers studying the web graph, market graphs, and online social networks have benefited from the application of network science and graph theory; see [**4, 12, 13**].

The study of complex networks has become an important topic of interdisciplinary research in the twenty-first century. Complex networks are found every where in nature and man-made systems, and complex networks arise in a wide range of settings, such as social, financial, and biological ones; see [**4**]. Nowadays, online social networks (OSNs) such as Facebook, Instagram, and Twitter have become increasingly popular, where nodes are user accounts and edges correspond to social interaction such as friendship or following. The popularity of OSNs has led to an increasing interest among researchers in the modeling and mining of real-world networks. To explain phenomena in complex networks, Watts and Strogatz introduced the small world property; see [**16**]. The small world property assumes low average distance and high clustering. Bonato et al. [**3**] introduced the Iterated Local Transitivity (ILT) model for OSNs and other complex networks that dynamically simulates many of their properties. The central idea of ILT model is *transitivity*: if u is a friend of v, and v is a friend of w, then u is a friend of w.

In recent years, network theory has been used to analyze many large financial data-sets that can be represented as graphs. Financial networks can be used to understand the dynamics of the market and the effects of globalization, and this analysis is especially important since the 2008 financial crisis. Pardalos et al. [**13**] introduced a model where the stocks are nodes and the edges are to be determined by price correlation between the stocks. This so-called *market graph* was shown to follow a power law degree distribution and satisfy the small world property. In the market graph, for a pair of nodes u and v, if the correlation coefficient $C_u(v)$ based on the price fluctuations of u and v is greater than or equal to a specified threshold θ, (where $\theta \in [-1, 1]$ is a real number), then there will be an edge between u and v.

Based on the presence of negative correlation in complex networks, we present a new, simplified deterministic model based on the principle of anti-transitivity: a non-friend of a non-friend is a friend. The Iterated Local Anti-Transitivity (ILAT) model dynamically simulates properties in many complex networks, such as densification power laws, constant diameter, and high local clustering. In social networks, negative correlation corresponds to enmity between agents in the network, while in stock market graphs, negative correlation corresponds to stocks that move in opposite directions with their prices.

1.2. Graph Theory

There are many circumstances in the real-world where networks arise. For example, the *nodes* known as fundamental unit of which the graphs are formed, could represent stocks with *edges*, the lines connecting nodes, formed between pairs of negatively correlated stocks. In the present section, we introduce graph theory and graph terminology; for additional background, see the books [**7, 17**].

Let V be the non-empty set of objects called *nodes* and E is a set of two-element subsets of V called *edges*. For simplicity, we refer to an edge $\{u, v\}$ by uv, and we say that u and v are *adjacent*. The ordered pair $G = (V, E)$ is called a *graph*.

We now give an example. Let $G = (V, E)$, where $V = \{a, b, c, d, 1, 2, 3, 4\}$ and

$$E = \{ab, bc, cd, ad, bd, a3, b4, c1, d2\}.$$

A drawing of the graph of G is shown in Figure 1.1.

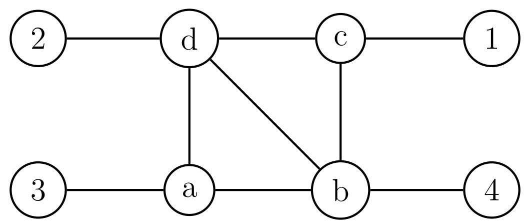

FIGURE 1.1. The graph $G = (V, E)$.

The sets of nodes and edges will be denoted by $V(G)$ and $E(G)$, respectively.

The *neighbourhood* of a node $u \in V(G)$ is the set of all nodes adjacent to u, and denoted by $N(u)$; that is, $N(u) = \{v \in V(G) : uv \in E(G)\}$. The *closed neighborhood* of u is written as $N[u]$ and expressed as $N[u] = N(u) \cup \{u\}$. The set of all nodes which are not adjacent to u (not including u) is denoted by $N^c(u)$, and is called the *non-neighbour set* of u.

The *degree* of node u is denoted by $\mathrm{deg_G}(u)$ or deg(u) and is the number of nodes adjacent to u; that is, the cardinality $|N(u)|$. If deg(u) $=$ 0, then the node u is called *isolated.* If deg(u) $= |V(G)| - 1$, then the node u is called *universal. The maximum degree* and *minimum degree* in a graph G is written as $\delta(G)$ and $\Delta(G)$, respectively. The following theorem is often refereed to as the *First Theorem of Graph Theory* (while the result is a part of folklore, we include its short proof for completeness).

Theorem 1. *If G is a graph, then we have that*

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

.

Proof. For each degree counted, we count each edge twice. $\square$

1.3. Special graphs

A *walk* in a graph $G = (V, E)$ from node u to node v is a sequence $W = (u = v_0, v_1, \ldots, v_l = v)$ if $v_i v_{i+1} \in E$ for $0 \leq i < l$. The *length* of a walk W is the number of nodes in W *minus 1* (that is, the number of edges). A walk is *closed* if $v_0 = v_l$. A *path* is a walk in which the internal nodes are distinct. The path of order n is denoted by P_n. Note that, the *order* of a graph is the cardinality of the number of nodes in the graph. For example, in Figure 1.1, $3 \to a \to b \to 4$ is a path.

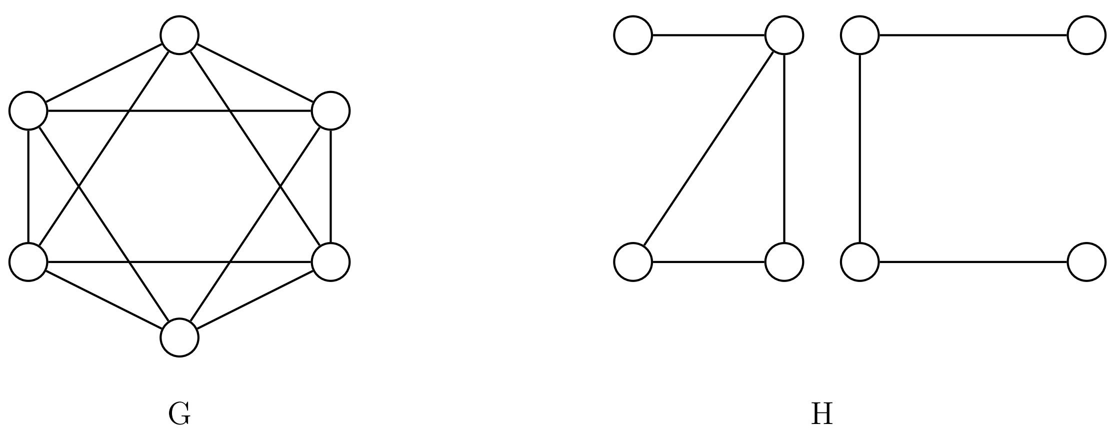

Figure 1.2. A connected graph G and a disconnected graph H.

If for all nodes $u, v \in V(G)$ there is a path whose endpoints are u and v, then G is called a *connected graph*; otherwise, the graph is said to be a *disconnected*. In Figure 1.2, G is connected and H is disconnected.

A graph G with a finite number of nodes is called *finite*. We only consider finite graphs in this thesis. The maximum number of possible edges in a simple graph with n nodes is $\binom{n}{2} = \frac{n(n-1)}{2}$. A graph is said to be a *complete* graph if for each pair of distinct nodes are adjacent. We denote the complete graph of n nodes by K_n. See Figure 1.3.

FIGURE 1.3. Complete graphs K_4 and K_6.

A *cycle* is a closed path of length at least 3. We use the notation C_n for a cycle of order n.

A *subgraph* H of a graph G satisfies $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. Hence, we can obtain subgraphs of a graph by deleting edges and nodes. In the following figure, H is the subgraph of G.

A subgraph H of a graph G is called an *induced subgraph* of G if all the edges between the nodes in H are exactly those in G. The subgraph induced by H can be written as $\langle S \rangle_G$. A set of nodes with no edges between them is called an *independent set*.

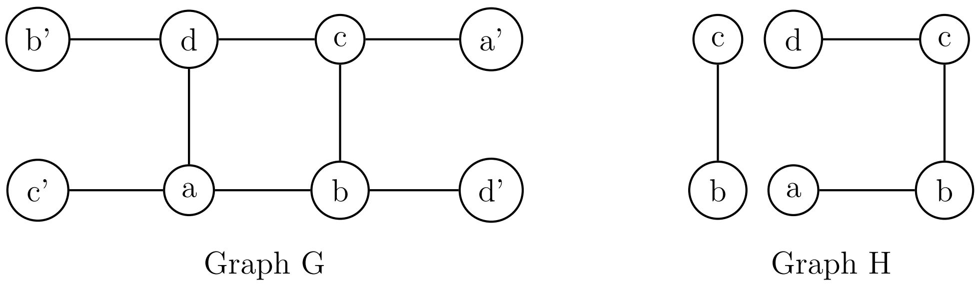

FIGURE 1.4. The subgraph H of the graph G.

A subgraph H of a graph G is called *spanning subgraph* if H and G have exactly the same node set; that is, $V(H) = V(G)$. Note that each induced subgraph is a subgraph, but the converse is false.

1.4. Isomorphism, distances, and adjacency matrices

Let G and H be two graphs. A *homomorphism* f between two graphs G and H is a function $f : V(G) \to V(H)$ such that, if for any nodes u and v, an edge $uv \in E(G)$ implies $f(u)f(v) \in E(H)$. An *embedding* from G to H is an injective homomorphism $f : V(G) \to V(H)$ with the property that $xy \in E(G)$ if and only if $f(x)f(y) \in E(H)$. An *isomorphism* is a bijective embedding.

If there is an isomorphism between two graphs, then we say that they are *isomorphic*. If G and H are isomorphic, then we write $G \cong H$. The graphs in Figure 1.5 are isomorphic.

In a connected graph G, the *distance* from node u to node v is the length of a shortest u-v path in G, and is denoted by $d(u, v)$ or $d_G(u, v)$. The *diameter* of the connected graph G is the supremum of all distances

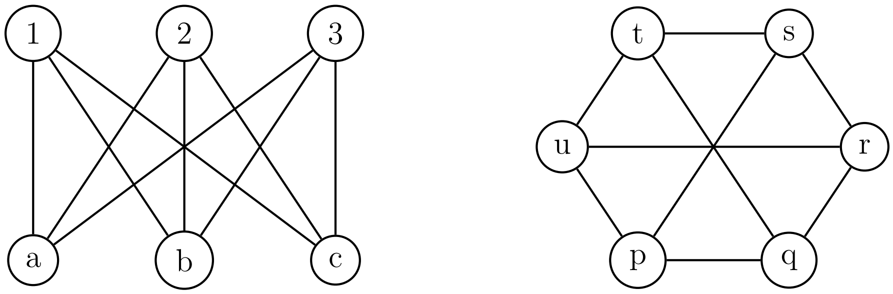

FIGURE 1.5. Isomorphic graphs.

between distinct pair of nodes and is denoted by diam(G). If the graph is disconnected, then the diameter of the graph is ∞.

In the following figure, the diameter of the graph G is 7.

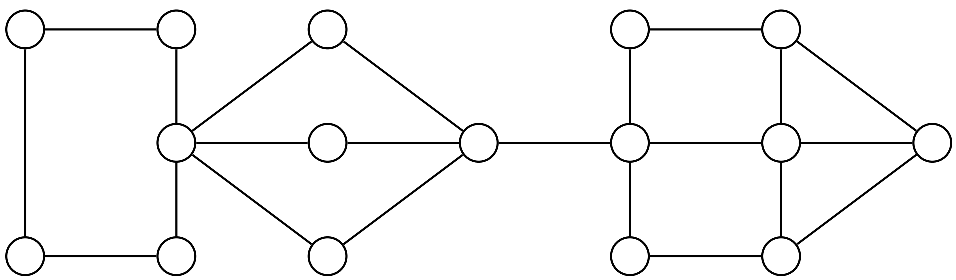

FIGURE 1.6. The graph G.

If G is a graph with n nodes, then the *adjacency matrix* $A(G)$ of G is the $n \times n$ binary matrix defined as follows:

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

Note that the adjacency matrix is symmetric with zeroes on the diagonal. We consider an example in Figure 1.7.

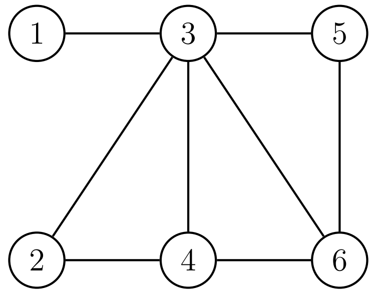

FIGURE 1.7. A graph H.

The graph H has adjacency matrix as follows.

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

1.5. Outline of the Thesis

In this chapter, we presented basic definitions, and notation from graph theory, together with a few fundamental properties and theorems in the field. We will present the ILAT model with examples and basic properties of the model in Chapter 2. We will discuss additional properties of the model in Chapter 3, where we will find the diameter, density, local clustering coefficient, and adjacency matrices of the graphs generated by the ILAT model. In Chapter 4, we focus on the domination numbers and cop numbers of ILAT graph. We will conclude the thesis in

Chapter 5 where we will summarize our results and have a brief discussion and also mention open problems and possible future directions.

CHAPTER 2

Iterated Local Anti-Transitivity Model

2.1. Introduction

In the present chapter, we introduce the *Iterated Local Anti-Transitivity* (or *ILAT*) model for social networks and other complex networks. In the ILAT model, we are given an initial graph, and new nodes will be repeatedly added over discrete time-steps by a process of anti-cloning. Anti-clones form an independent set and are adjacent to the nodes in the complement of the closed neighborhood of their host node. In each of the duplication models for protein-protein interactions, the copying model, and in the ILT model there was a notion of cloning nodes; see [**3, 4**].

The ILAT model is somewhat analogous to the ILT model as studied in [**3**]. Observe that the ILAT model displays a complex, dependency structure of anti-cloned nodes over time. Like the ILT model (and unlike the duplication and copying models, as described in [**2**]), it is deterministic. The ILAT model may be considered as a simplified version of the duplication model, whereby all nodes are anti-cloned in a given time-step, rather than nodes being duplicated one-by-one via probabilistic rules. Unlike the ILT model where a node and its clone are adjacent, in the ILAT model, a node and its anti-clone will not be adjacent.

2.2. Formulation of the model

The ILAT model generates simple, finite and undirected graphs. There is a single parameter which is the initial graph G_0, which is a graph. For a non-negative integer t, the graph G_t represents graph at time-step t. Suppose that the graph G_t has been defined for a fixed time $t \geq 0$. To form the graph G_{t+1}, for each node $v \in V(G_t)$, we add its *anti-clone* v' such that v' is adjacent to the *non-neighbour set* of v: that is, the nodes in $V(G_t)$ except v and its neighbour set in G_t.

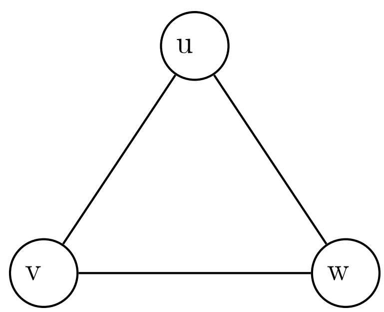

FIGURE 2.1. The initial graph G_0.

For example, let G_0 be K_3 and set $V(G_0) = \{u, v, w\}$. See Figure 2.1. At time $t = 1$, we will have the set of nodes $V(G_1) = \{u, v, w, u', v', w', \}$ by anti-cloning u, v, w, and the anti-clones will not be adjacent to any node in u, v, w. That is, no edges will be created in this step because the complement of the neighborhood of each node is empty. See Figure 2.2. Similarly, in the next time-step $t = 2$ the anti-clones of each nodes in $V(G_1)$ will be created. We will have the set of nodes $V(G_2) = \{u, v, w; u', v', w'; u'', v'', w''(u')', (v')', (w')'\}$ in time-step $t = 2$. In this time-step, each of the nodes in $\{u'', v'', w''\}$ are adjacent to each of the nodes in $\{u', v', w'\}$, and each of the nodes in $\{(u')', (v')', (w')'\}$ are adjacent to each of the nodes in $\{u, v, w\}$.

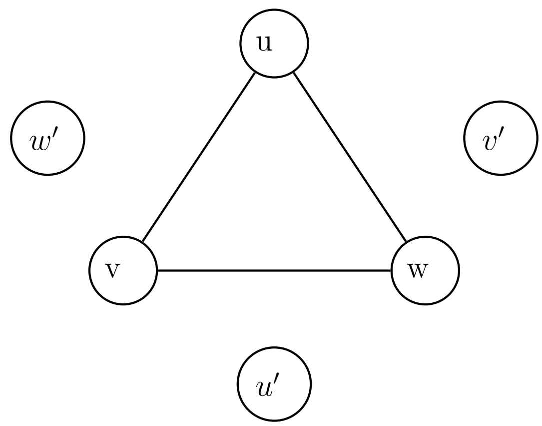

FIGURE 2.2. The graph G_1 at $t = 1$.

Next, we add new nodes by taking anti-clones of each nodes in time-step $t = 0$ to create a graph in time-step $t = 1$. Note that the nodes created in this time are isolated. In time-step $t = 2$, we will have twelve nodes. In this time-step, there are no isolated nodes. Anti-clones are created from the nodes $\{u, v, w\}$ (introduced at $t = 0$) will be adjacent to $\{u', v', w'\}$. Anti-clones created at $t = 2$ from the nodes $\{u', v', w'\}$ (introduced in time $t = 1$) will be adjacent to the nodes $\{u, v, w\}$ created at $t = 0$. See Figure 2.3. We also provide figures of the time-steps $1 \leq t \leq 4$ with initial graph C_4.

2.3. Basic properties of the ILAT model

Note that the set of new nodes created at $t + 1$ from the set of nodes of the graph G_t is an independent set with cardinality $|V(G_t)|$. Since we create a new graph by taking anti-clones of each nodes from the given

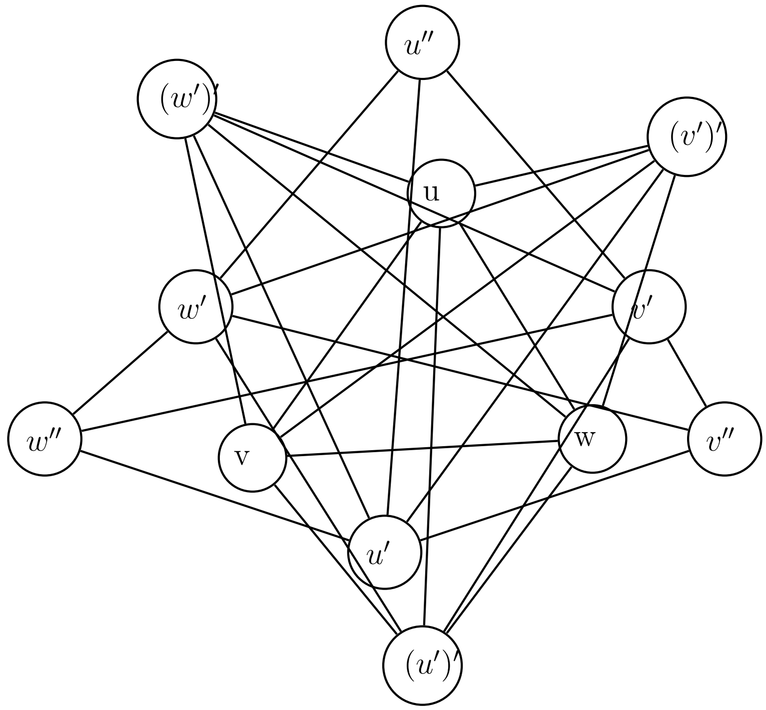

FIGURE 2.3. The graph G_2 at $t = 2$.

graph, we double the number of nodes at each time-step. Hence, if the number of nodes in G_t is denoted by n_t, then we observe that $n_t = 2^t n_0$, where n_0 is the number of nodes in starting graph G_0.

We let $\mathrm{deg_t(u)}$ represent the degree of node u in the graph G_t at t. We notice in Figure 2.2 above that even if we have a connected initial graph, then the graph in later time-steps may have isolated nodes. Our next result shows that if we let the graph process run for two time-steps, then there are no isolated nodes nor universal nodes.

THEOREM 2. *Let $t \geq 0$ be an integer.*

(1) *For $t \geq 1$, the graph G_{t+1} has no universal nodes.*

(2) *For $t \geq 2$, the graph G_{t+1} has no isolated nodes.*

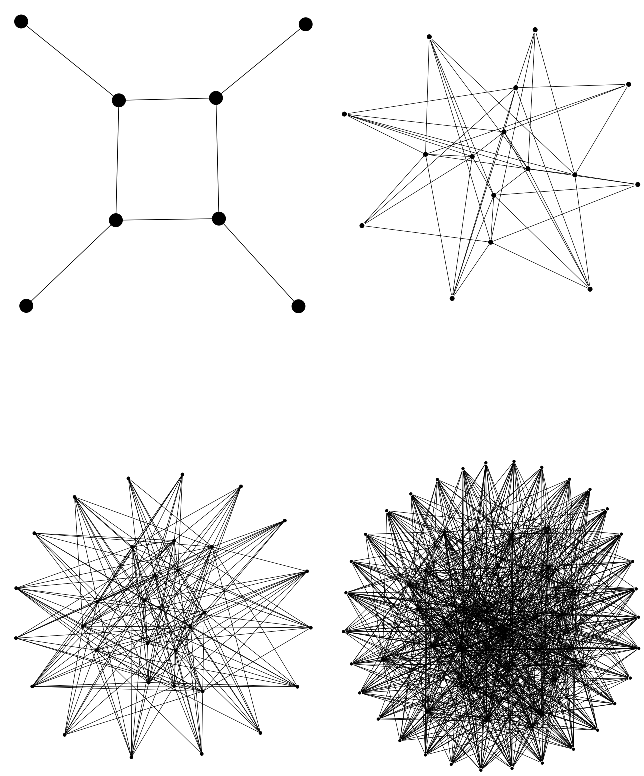

FIGURE 2.4. An example of the first four time-steps of the ILAT model, where the initial graph is the four-cycle C_4.

PROOF. For item (1), note that an existing node u is not adjacent to its anti-clone u' and so cannot be universal. Analogously, the anti-clone u' is not adjacent to u and so cannot be universal.

For item (2), suppose for a contradiction that G_{t+1} contains an isolated node v and $t \geq 2$. There are two cases. First, suppose that v is an anti-clone w' added at time $t+1$. Then w would be universal in G_t (where $t \geq 1$) and that contradicts item (1). Second, suppose v was an existing node from G_t. Then v is not adjacent to any node in G_t, and since $t \geq 2$, there is at least one node of G_t, say z, distinct from v. In that case, z' is adjacent to v, which is a contradiction. $\square$

Note that the bounds on t in Theorem 2 are sharp. For example, let G_0 be K_3. Then G_0 has universal nodes, and G_1 contains three isolated nodes (as the anti-clone of each vertex from time $t = 0$ is isolated).

We now derive recursive formulas for the degrees of nodes in graphs generated by the ILAT model. Let the degree of a node x at time t be denoted $\deg_t(x)$. We define the *co-degree* of x at time t as ${\deg^c}_t(x) = n_t - \deg_t(x) - 1$. The co-degree of x counts the number of nodes not adjacent to x.

LEMMA 3. For $t \geq 1$, and $u \in G_t$ we have the following

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

(2) $\deg_{t+1}(u') = n_t - \deg_t(u) - 1$.

PROOF. For item (1), we note that the node u retains its edges from time-step t to $t + 1$. Further, u acquires new edges to anti-clones w', where w is not adjacent to u in G_t. Hence, there is an edge from u to a new node precisely when there is a non-edge to an existing node. We therefore, derive that

$$\begin{aligned}\deg_{t+1}(u) &= \deg_t(u) + {\deg^c}_t(u) \\ &= \deg_t(u) + n_t - \deg_t(x) - 1 \\ &= n_t - 1,\end{aligned}$$

and the result follows.

For item (2), an anti-clone u' is adjacent exactly to the non-neighbours of u. In particular, $\deg_{t+1}(u')$ equals ${\deg^c}_t(u)$. The result now follows from the definition of the co-degree of a node. □

Note that by Lemma 3 we have that for all $x \in V(G_t)$,

$$\deg_t(x) \leq n_{t-1} - 1.$$

Since $n_{t-1} = \frac{1}{2}n_t$, we derive that

$$\deg_t(x) \leq \frac{1}{2}n_t - 1.$$

We define $\mathrm{Vol}_t = \sum_{v \in V(G_t)} \deg v$. The quantity Vol_t is referred to as the *volume* of the graph at time t. Note that $\mathrm{Vol}_t = 2e_t$, by th first theorem of the graph theory. The following lemma sets up a recursive formula for the volume, and hence, the number of edges.

LEMMA 4. 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

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

where the third equality follows by Lemma 3.

Next, note that

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

where the second equality follows by the previous derivation in this lemma for the volume. $\square$

Next, we prove that the ILAT model generates graphs which densify over time; that is, the average degree of the graphs tend to infinity. One interpretation is that in networks where anti-transitivity is pervasive, many alliances form in the network over time.

Theorem 5. *In the ILAT model,* $\frac{e_t}{n_t} \to \infty$

Proof. By Lemma 4, we have that

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

Hence, by recursion, we find that

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

where

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

and where

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

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 infinity. 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 the result follows. □

From this discussion, we can obtain a limiting density as $t \to \infty$. Let D_t be the density of G_t; that is, $D_t = \frac{e_t}{\binom{n_t}{2}}$. The ILAT model generates quite dense graphs. Let $f(t)$ and $g(t)$ be two non-negative integer valued functions. We use the asymptotic notation $f(t) \sim g(t)$ for $f(t) = (1 + o(1))g(t)$.

Corollary 6. *As $t \to \infty$, we have that $D_t \to 2/5$.*

Proof. Note that $D_t = 2\frac{e_t}{n_t(n_t-1)}$. From the proof of Theorem 5 and the identity $n_t = 2^t n_0$, we then have that

$$\begin{aligned} D_t &= \frac{2^{2t}(n_0)^2}{n_t(n_t-1)} \left(\frac{1}{5}\right)\left(1-\left(-\frac{1}{4}^{t-1}\right)\right)(1-o(1)) \\ &\sim 2/5, \end{aligned}$$

and the proof follows. $\square$

Below is a table which shows the density of ILAT graph whose starting graph is $G_0 \cong K_3$.

Graphs	Total no of nodes	Total no of edges	Graph density
G_0	3	3	1
G_1	6	3	0.2
G_2	12	27	0.409
G_3	24	105	0.380
G_4	48	447	0.396

CHAPTER 3

Additional properties of the ILAT graph

3.1. Introduction

We discussed some of the basic properties of the ILAT model in Chapter 2 and now we discuss some additional properties. Our focus will be on distances, clustering, and spectral properties of graphs generated by ILAT.

The distance between two nodes in a graph is a simple but surprisingly useful notion. Many problems in data communication can often be related to a few key invariants of which the diameter is an important one. In this chapter, we will show that for $t \geq 2$, any two nodes that are not newly created are at most distance 2 apart. We will next show that for the average distance in ILAT graphs, all but a negligible number of pairs of nodes have distance at most 2. We will then find that the distances within graphs generated by ILAT at $t \geq 2$ become very small; in particular, we will show they have diameter 3 in Theorem 7.

3.2. Diameter of ILAT graphs

The diameter of a graph is the length of shortest path between the most distanced nodes. In ILAT graphs, distances become small in succeeding time steps, even if the initial graph has high diameter. More precisely, we will show in Theorem 7 that the diameter with in the ILAT graph is 3 at time step $t = 2$, regardless of the diameter of the starting graph G_0.

As an illustrative example, we can consider the graph $G_0 = C_4$ with diameter $\mathrm{diam}(G_0) = 2$. At $t = 1$ we derive that $\mathrm{diam}(G_1) = 4$. But at $t = 2$, the diameter of the graph G_2 reduces to 3. See Figure 3.2.

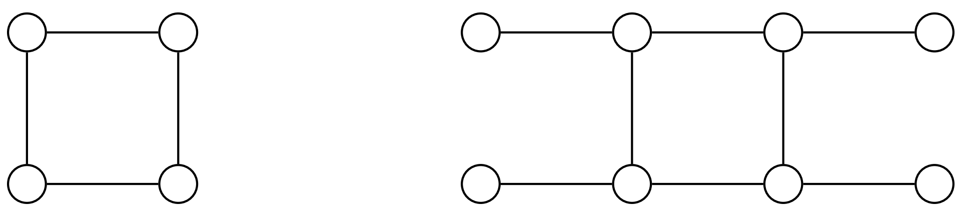

FIGURE 3.1. The graphs G_0 of diameter 2 and G_1 of diameter 4.

For an other example we take K_3 as the initial graph G_0. Since the graph is disconnected at $t = 1$, the diameter at that time-step is infinite. At time step $t = 2$, however, the graph G_2 is connected and will have $\mathrm{diam}(G_2) = 3$.

Our main theorem of this section is the following.

THEOREM 7. *If* $t \geq 2$, *then* $\mathrm{diam}(G_t) = 3$.

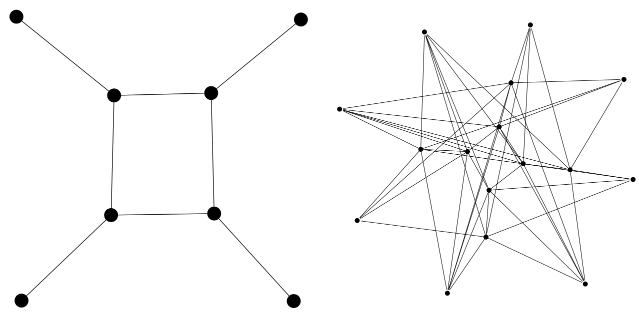

FIGURE 3.2. The graphs G_1 of diameter 4 and G_2 of diameter 3, where the initial graph is C_4 of diameter 2.

We will prove Theorem 7 after proving the following lemmas. The first derives a lower bound.

LEMMA 8. If $t \geq 1$, then $\text{diam}(G_t) \geq 3$, for all initial graphs.

PROOF. Let us consider a node u at $t-1$ or older and its anti-clone u' at t. Since u and u' are not adjacent and have no common neighbors, we have that $d(x, x') \geq 3$. This proves the lemma. $\square$

We next consider nodes that are not newly created.

LEMMA 9. If $t \geq 2$, then the distance between two nodes that are not newly created is at most 2.

PROOF. Let us consider two nodes u and v in G_t that both existed in G_{t-1}, and were born at time $t-1$ or earlier. By the discussion after Lemma 3, the degree of nodes at $t-1$ are bounded above by $\frac{n_t}{4} - 1$. Hence, there exists another node w that also existed at $t-1$ or earlier that is not adjacent to either of u or v. It follows that the anti-clone w' of w is adjacent to both of nodes u and v, and the proof follows. $\square$

We last consider newly created nodes.

LEMMA 10. Every pair of nodes that were both newly created are at most distance 2 apart.

PROOF. Let $u', v' \in G_t$ be two nodes that both are newly created at t by taking anti-clones of some nodes u, v created at time $t-1$ or earlier. Since the degree of a node at $t-1$ is bounded above by $\frac{n_t}{4} - 1$, there is another node w that also existed at time $t-1$ or earlier that is not adjacent to either u or v. These properties imply that the node w is adjacent to both u' and v', and so $d(u', v') \leq 2$. $\square$

We now complete the proof of the theorem.

PROOF OF THEOREM 7. The lower bound follows by Lemma 8. The cases for the upper bound for not newly created nodes follows by Lemma 9, and for both newly created nodes by Lemma 10. The final case is that one node is newly created and the other is not. If $t \geq 2$, then every newly created node has a neighbour that is not newly created; the analogous statement holds for nodes that are not newly created. Therefore, any such pair can be connected by a path of length at most 3. $\square$

3.3. Clustering coefficients of ILAT graphs

The *local clustering coefficient* for a node $v \in G_t$, written $c_t(v)$, is the proportion of edges between the nodes within its neighbourhood divided by the number of all possible edges between them. To be more precise, for a set S of nodes its *density* is the number of edges in the subgraph induced by S, divided by $\binom{|S|}{2}$. Thus, the local clustering coefficient of v in G_t is the density of its neighbour set. More explicitly,

$$c_t(v) = \frac{2|uv : u, v \in N_t(v), uv \in E_t|}{|N_t(v)|(|N_t(v)| - 1)}.$$

The *clustering coefficient* of a graph G_t, written $c(G_t)$, is the average of its local clustering coefficients. Hence, we have that

$$c(G_t) = \frac{1}{n_t} \sum_{v \in V(G_t)} c_t(v).$$

Older nodes in graphs generated by the ILAT model exhibit high local clustering in the following theorem, as proved first in [**4**]. We omit the proof of this theorem, which may be found in [**4**].

THEOREM 11 ([**4**]). *For any node $v \in V(G_t)$ created at $k < t$, if $\lim_{t\to\infty} c_t(v)$ exists, then we have that $\lim_{t\to\infty} c_t(v) = \frac{2}{5}$.*

Note that we make the assumption in Theorem 11 that $\lim_{t\to\infty} c_t(v)$ exists, and we cannot remove this restriction at the time of writing this thesis.

3.4. Adjacency matrices of ILAT graphs

If G is a graph with n nodes, recall that the *adjacency matrix* $A(G)$ is the $n \times n$ matrix defined as follows:

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

Define J_t to be the *all-ones matrix* of order n_t, I_t to be the *identity matrix* of order n_t, and O_t is the *zero matrix* of order n_t. In ILAT graphs, if A_t is the adjacency matrix of G_t, then the adjacency matrix of G_{t+1} is given by:

$$A_{t+1} = \begin{pmatrix} A_t & J_t - I_t - A_t \\ J_t - I_t - A_t & O_t \end{pmatrix}.$$

Since the graph is undirected the adjacency matrix is symmetric: the *ith* row, *jth* column entry is 1 if and only if the *jth* row and *ith* column entry is 1, and all the entries in the major diagonal is zero. The size of adjacency matrix is equal to the number of nodes in the graph. If n_t is the number of nodes in graph G_t, then the order of adjacency matrix A_t is $n_t \times n_t$. Hence, the order of adjacency matrix A_{t+1} is $2n_t \times 2n_t$.

For example, let us consider C_4 as the initial graph whose adjacency matrix of order 4×4 is:

$$A_0 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Then the adjacency matrix of the ILAT graph at $t = 1$ is the matrix of order 8×8 given by:

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The adjacency matrix of the graph at $t = 2$ is the matrix of order 16×16 given by:

$$A_2 = \begin{pmatrix} 0&1&0&1&0&0&1&0&0&0&1&0&1&1&0&1 \\ 1&0&1&0&0&0&0&0&0&0&0&1&1&1&1&0 \\ 0&1&0&1&1&0&0&0&1&0&0&0&0&1&1&1 \\ 1&0&1&0&0&1&0&0&0&1&0&0&1&0&1&1 \\ 0&0&1&0&0&0&0&0&1&1&0&1&0&1&1&1 \\ 0&0&0&1&0&0&0&0&1&1&1&0&1&0&1&1 \\ 1&0&0&0&0&0&0&0&0&1&1&1&1&1&0&1 \\ 0&1&0&0&0&0&0&0&1&0&1&1&1&1&1&0 \\ 0&0&1&0&1&1&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&1&1&1&0&0&0&0&0&0&0&0 \\ 0&1&0&0&1&0&1&1&0&0&0&0&0&0&0&0 \\ 1&1&0&1&0&1&1&1&0&0&0&0&0&0&0&0 \\ 1&1&1&0&1&0&1&1&0&0&0&0&0&0&0&0 \\ 0&1&1&1&1&1&0&1&0&0&0&0&0&0&0&0 \\ 1&0&1&1&1&1&1&0&0&0&0&0&0&0&0&0 \end{pmatrix}.$$

CHAPTER 4

Domination and Cop Numbers of ILAT graphs

4.1. Introduction

In this chapter, we determine the domination numbers and cop numbers of ILAT graphs. The study of domination in graphs came about partially as a result of the study of games and recreational mathematics. In particular, mathematicians studied how chess pieces of a particular type could be placed on a chessboard in such a way that they would attack, or dominate, every square on the board. Although the formal mathematical study of domination in graphs began around 1960, there are some references to domination-related problems even one hundred years prior. In 1862, De Jaenisch attempted to determine the minimum number of queens required to cover an $n \times n$ chess board. See [**9**] for further background.

The study of domination was further developed in the late 1950's and 1960's, beginning with the work of Berge [**1**]. Berge introduced the *coefficient of external stability,* which is now known as the domination number. In a graph G, $S \subseteq V(G)$ is said to be a *dominating set* if every node $v \in V(G)$ is either in S or adjacent to a node in S. The *domination* *number* of a graph G, written $\gamma(G)$, is the minimum cardinality of the dominating set of G. Note that $\gamma(G)$ is well-defined since $\gamma(G) \leq |V(G)|$.

The game of Cops and Robbers, introduced independently by Nowakowski and Winkler [**11**] and Quilliot [**14, 15**] over twenty-five years ago. Researchers have studied this game, as well as many variants of the original game; see [**5**] for additional history and background on the game. The game of *Cops and Robbers* is an example of a vertex-pursuit game played on a graph. The game is played as follows. We have players, the *robber* and the *cops*. It is permissible for the cop player to have a set of cops at their disposal. First, the cops are placed on the nodes of the graph. Then the robber is placed on a node. The two players play in alternating turns. If we have more than one cop, then the cop player moves a subset of the cops. In any given round, either player may pass by remaining on their node. The cop wins if he captures the robber, and the robber wins if robber can avoid capture indefinitely.

The minimum number of cops needed to win the game on G is the *cop number* of G, and is written $c(G)$. Note that $c(G)$ is well-defined since it is bounded above by $|V(G)|$ (and it is also bounded above by $\gamma(G)$).

4.2. Domination in ILAT graphs

We now give the detail description of the domination number of the ILAT graphs. We begin with an example as described in the figure.

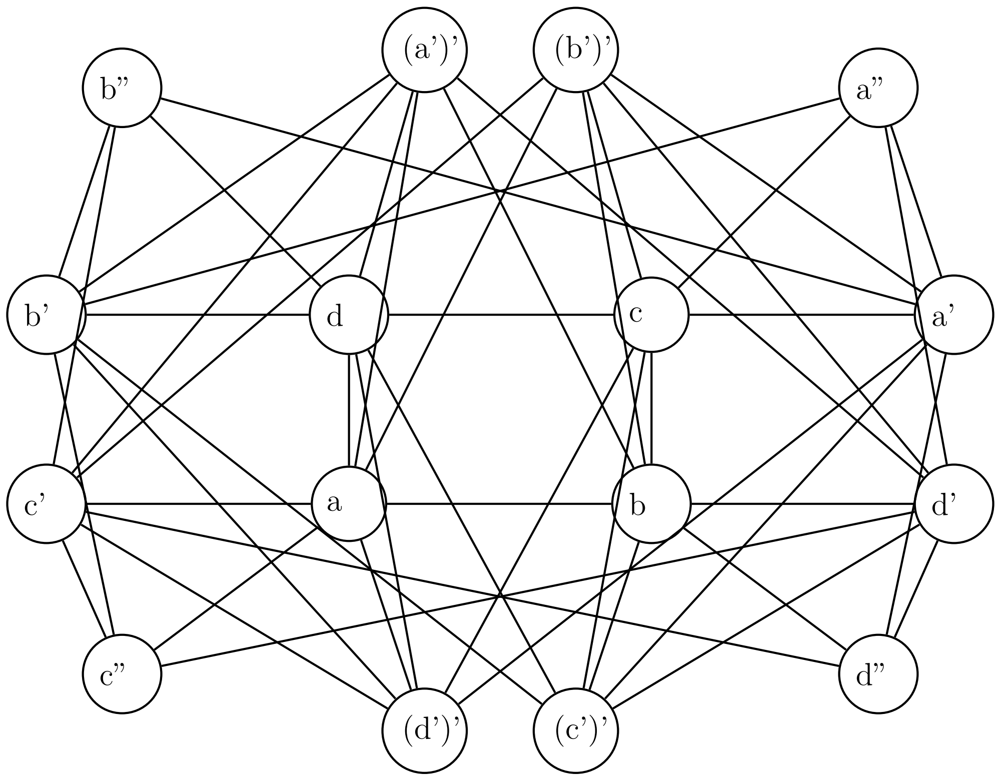

FIGURE 4.1. The ILAT graph G_2.

The ILAT graph G_2 shown in the Figure 4.1 had initial graph $G_0 = C_4$ with nodes $\{a, b, c, d\}$. Observe that the set of nodes of $V(G_2)$ can be written as

$$\{a, b, c, d; a', b', c', d'; (a')', (b')', (c')', (d')'; a'', b'', c'', d''\},$$

where a', b', c', d' are the anti-clones created at $t = 1$, and a'', b'', c'', d'' are the anti-clones created at $t = 2$. Similarly, the anti-clones $(a')', (b')', (c')', (d')'$ are the anti-clones created at $t = 2$ from the nodes a', b', c', d' created at $t = 1$. Now it can be seen from Figure 4.1 that either of the sets of nodes $\{a, a', (a')'\}$, $\{a, a', a''\}$, $\{b, b', (b')'\}$, $\{b, b', b''\}$, $\{c, c', (c')'\}$ or $\{c, c', c''\}$,

$\{d, d', (d')'\}$, or $\{d, d', d''\}$ forms a dominating set. For example, in the sequence $a, a', (a')'$, the node a is adjacent to the anti-clones created at $t = 2$ from the nodes created at $t = 1$, the node a' is adjacent to the anti-clones created at $t = 2$ from the nodes at $t = 0$ and the node $(a')'$ is adjacent to the anti-clones created at $t = 1$ from the nodes at $t = 0$. Hence, the domination number of the graph G_2 is $\gamma(G_2) \leq 3$.

We first prove in Lemma 12 that the domination number of the ILAT graph $\gamma(G_t) > 2$ for a sufficiently large t, and then we prove $\gamma(G_t) = 3$ in Theorem 13.

LEMMA 12. If $t \geq 3$, then $\gamma(G_t) > 2$.

PROOF. Let $u, v \in V(G_t)$ be two nodes. If both u, v existed at time $t - 1$, then by the discussion after Lemma 3, the degree of nodes at $t - 1$ are bounded above by $\frac{n_t}{4} - 1$. Hence, there exists another node w that also existed at $t - 1$ or earlier that is not adjacent to either of u or v.

If u existed at time $t - 1$ and v was newly created at time t, then u' is not adjacent to either u or v. The case when u is newly created and v was an existing node is analogous.

In all cases, the nodes u, v do not dominate G_t. Hence, $\gamma(G_t) > 2$. $\square$

We now come to the main theorem of this section.

THEOREM 13. *If* $t \geq 3$, *then* $\gamma(G_t) = 3$.

PROOF. We show that the three nodes $u, u', (u')'$ form a dominating set, where u is the node created at $t-2$ or older, u' the anti-clone of u at $t-1$ and $(u')'$ is that of u' created at t. Every node z not equalling $u, u', (u')'$ is either adjacent to u', adjacent to $(u')'$, or is a node created at time $t-1$ that is not adjacent to u'. In the latter case, z must be adjacent to u. This shows that $\gamma(G_t) \leq 3$. The lower bound follows from Lemma 12. □

4.3. Cop numbers of ILAT graphs

We will prove in Theorem 14 that ILAT graphs eventually have cop number at most 2, regardless of the starting graph (which could have arbitrary large cop number). To further illustrate the game, we begin with an example where we have a single cop as depicted in the following figures.Suppose that the cop starts from the leftmost node first, and suppose that the robber will take place the position as far away from the cop as possible.

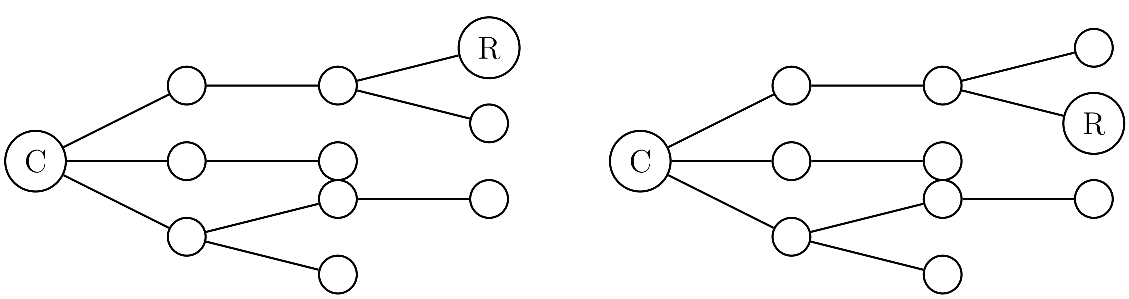

FIGURE 4.2. Cops and Robbers: rounds 1 and 2.

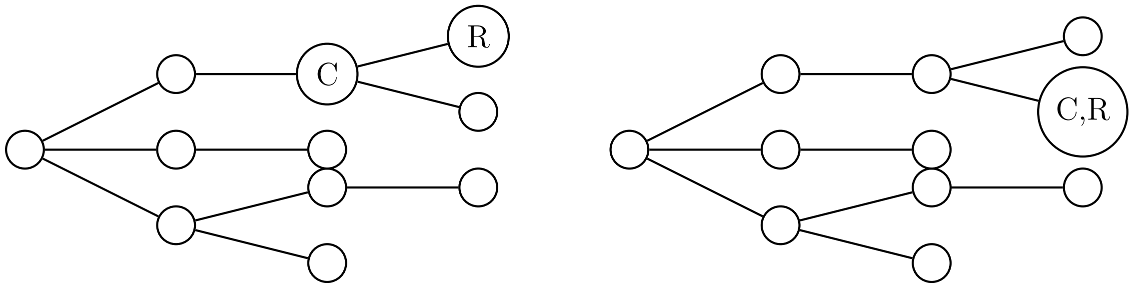

FIGURE 4.3. Cops and Robbers: rounds 3 and 4.

In the second round, the cop moves first as closer to the robber. If the robber does not pass, then they will be closer to the cop. Hence, the robber will stay put and await capture.

We now arrive at the main theorem of this section.

THEOREM 14. *If $t \geq 2$, then $c(G_t) \leq 2$.*

PROOF. We consider a partition of the set of nodes in the ILAT graph G_t into sets A, B, C and D. Let A consist of the nodes created at $t-2$ or earlier, let B be their anti-clones at $t-1$ and C and D are anti-clones of A and B respectively, created at t. We use the notation x'' for nodes in C and D, where x is either in A or B.

Let C_1,C_2 be the two cops. In the first step, place one cop C_1 on node $u \in A$ and other cop C_2 on node $u'' \in C$. Each node of G_{t-1} is adjacent to one of u or u'', so the robber must choose their initial vertex v'' in one of C or D.

As $d_t(u'', v'') \leq 2$ by Lemma 10, there exists at least one common neighbour node say $w \in A \cup B$ to u'' and v''. Note that u is not adjacent to w, and so u is adjacent to w''. In the next round, C_2 moves to w and C_1 moves to w'', which then forces the robber to move to $A \cup B$ (as $C \cup D$ is an independent set). But then every vertex of $A \cup B$ is joined to at least one of w or w'' and the robber is captured. $\square$

CHAPTER 5

Conclusion and Future Work

5.1. Conclusion

Anti-transitivity is one generative process in the formation of ties in social networks and other complex networks. In this thesis, we introduced a simplified, evolutionary model for anti-transitivity in complex networks called the Iterated Local Anti-Transitivity (ILAT) model.

In Chapter 2, we presented a precise definition of the ILAT model and examined its basic properties, such as the number of edges and densification in the graphs it generates. We also derived the density of ILAT graphs tending to be 0.4, and considered the degrees of nodes of ILAT graphs. In Chapter 3, we proved that ILAT graphs have diameter 3 after sufficiently large time-steps, regardless of the diameter of the initial graph. We considered the local clustering coefficient of the older nodes in ILAT graphs, which exhibits significant local clustering over time. We also computed the adjacency matrix of the ILAT graphs over time.

We showed that the ILAT graphs have small dominating sets and low cop number in Chapter 4. We have proved in this chapter that the domination number $\gamma(G_t)$ of the ILAT graph G_t is equal to 3 for $t \geq 3$, regardless of its initial graph. We have also shown that the ILAT graphs have cop number at most 2 for sufficiently large time-steps.

5.2. Future work and open problems

Theoretical results presented here for the ILAT model are suggestive of several emergent properties in networks where anti-transitivity governs link formation. For instance, the presence of 3-element dominating sets suggest the emergence of small but important alliances. Such small dominating sets may emerge naturally in real-world networks which are highly anti-transitive, owing to a high number of alliances against common adversaries. The high density, low diameter, and high local clustering of graphs generated by the ILAT model suggest a network structure where there are tight-knit alliances formed against mutual enemies. For instance, we may find such network structures when studying conflict between nation states, rival companies in business, or even in networks of gangs. It would be interesting to apply the ILAT model to such real-world networked data.

Besides applications of the ILAT model, there remain a number of theoretical questions about the graphs it generates.

(1) While we found an asymptotic expression for the local clustering coefficient of older nodes in the ILAT model, an open problem remains to compute the clustering coefficient for ILAT graphs.

(2) Another question is to determine the induced subgraph structure of ILAT graphs. A characterization of the induced subgraphs of ILAT graphs (that is, to determine its *age*) remains open. For example, do all finite trees appear as induced subgraphs of ILAT graphs?

(3) For $t \geq 2$, is it ever the case that the graphs G_t have cop number 1?

(4) We introduce large independent sets at each time-step of the model, which may be less realistic and impact the value of the clustering coefficient. One approach to this issue is to form a random, dense subgraph on the anti-clones. Such a randomized ILAT model would be interesting to study in future work.

Appendix A - Code

We used the following code in MATLAB to find the adjacency matrix of ILAT graphs We then used these matrices to draw the graphs in Gephi.

```
A = [0,1,0,1; 1,0,1,0; 0,1,0,1; 1,0,1,0];
e = [];
for i = 1:2
j = size(A,1);

J = ones(j);
I = eye(j);
O = zeros(j);
B = J - I - A;
A = [A,B;B,O]
[V D] = eig(A);
V(:,j);
end
ab = 0;
for p = 1:size(A,1)
for j = 1:size(A,1)
```

```
if A(p,j) == 1
ab = ab + 1;
end
end
end
ab = ab/2
M =zeros(ab,3);
z = 1;
for p = 1:size(A,1)
for j = p:size(A,1)
if A(p,j) == 1
M(z,1) = p
M(z,2) = j
M(z,3) = 1
z = z+1
end
end
end
csvwrite('csvtest.csv',M)
```

Bibliography

[1] C. Berge, *Theory of Graphs and its Applications*, Methuen, London, 1962.

[2] A. Bonato, *A Course on the Web Graph*, American Mathematical Society Graduate Studies Series in Mathematics, Providence, Rhode Island, 2008.

[3] A. Bonato, N. Hadi, P. Prałat, C. Wang, Models of online social networks, *Internet Mathematics* **3** (2010) 285–313.

[4] A. Bonato, E. Infeld, H. Pokhrel, P. Prałat, Common adverseries from alliances: modeling complex network via anti-transitivity, In: *Proceedings of the 14th Workshop on Algorithms and Models for the Web Graph (WAW 2017)*, 2017.

[5] A. Bonato, R.J. Nowakowski, *The Game of Cops and Robbers on Graphs*, American Mathematical Society, Providence, Rhode Island, 2011.

[6] P. Carrington, J. Scott, *The SAGE handbook of social network analysis*, Sage Publications Ltd, 2011.

[7] R. Diestel, *Graph theory*, Springer-Verlag, New York, 2000.

[8] L. Freeman, Centrality in social networks: conceptual clarification, Social Networks **1** (1979) 215–239.

[9] T.W. Haynes, S.T. Hedetniemi, P.J. Slater, *Fundamentals of Domination in Graphs*, CRC Press, 1998.

[10] J. Moreno, *Who Shall Survive?: foundations of sociometry, group psychotherapy, and socio-drama*, Beacon House, 1978.

[11] R.J. Nowakowski, P. Winkler, Vertex-to-vertex pursuit in a graph, Discrete Mathematics 43 (1983) 235–239.

[12] R. Orendovici, Social network analysis and simulation of the development of adversarial networks, In: *Proceedings of BRIMS*, 2012.

[13] P. Pardalos, V. Boginski, S. Butenko, Statistical analysis of financial networks, *Computational Statistics and Data Analysis*, **48** (2005) 431–443.

[14] A. Quilliot, Jeux et pointes fixes sur les graphes, *Thèse de 3ème cycle*, Université de Paris VI, 1978, 131–145.

[15] A. Quilliot, *Problèmes de jeux, de point Fixe, de connectivité et de represésentation sur des graphes, des ensembles ordonnés et des hypergraphes*, Thèse d'Etat, Université de Paris VI, 1983, 131–145.

[16] D. Watts, S. Strogatz, Collective dynamics of small-world networks, *Nature* **393** (1998) 440–442.

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