DISTANCE-PRESERVING GRAPHS

By

Ronald Nussbaum

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Computer Science – Doctor of Philosophy

2014

ABSTRACT

DISTANCE-PRESERVING GRAPHS

By

Ronald Nussbaum

Let G be a simple graph on n vertices, where $d_G(u, v)$ denotes the distance between vertices u and v in G. An induced subgraph H of G is *isometric* if $d_H(u, v) = d_G(u, v)$ for all $u, v \in V(H)$. We say that G is a *distance-preserving graph* if G contains at least one isometric subgraph of order k for every k where $1 \leq k \leq n$.

A number of sufficient conditions exist for a graph to be distance-preserving. We show that all hypercubes and graphs with $\delta(G) \geq \frac{2n}{3} - 1$ are distance-preserving. Towards this end, we carefully examine the role of "forbidden" subgraphs. We discuss our observations, and provide some conjectures which we computationally verified for small values of n. We say that a distance-preserving graph is *sequentially distance-preserving* if each subgraph in the set of isometric subgraphs is a superset of the previous one, and consider this special case as well.

There are a number of questions involving the construction of distance-preserving graphs. We show that it is always possible to add an edge to a non-complete sequentially distance-preserving graph such that the augmented graph is still sequentially distance-preserving. We further conjecture that the same is true of all distance-preserving graphs. We discuss our observations on making non-distance-preserving graphs into distance preserving ones via adding edges. We show methods for constructing regular distance-preserving graphs, and consider constructing distance-preserving graphs for arbitrary degree sequences. As before, all conjectures here have been computationally verified for small values of n.

To my friends and loved ones.

ACKNOWLEDGEMENTS

I am extremely grateful to everyone who helped me during my Ph.D. journey. Without their guidance and support, such an undertaking would not have been possible. My deepest thanks go to my advisor, Dr. Abdol-Hossein Esfahanian, and the other members of my Ph.D. committee: Dr. Eric Torng, Dr. Pang-Ning Tan, and Dr. Erik Goodman. All of these individuals have been excellent teachers and mentors since I first began my M.S. studies. Dr. Esfahanian was a wonderful advisor, keeping me motivated and focused throughout my Ph.D. studies. A number of colleagues assisted with the study of distance-preserving graphs in recent years. I want to thank Tobias Ludwig, Dennis Ross, Dr. Bruce Sagan, and Emad Zahedi for their ideas and contributions to joint work.

Along with my research, I owe a great debt to the faculty of the departments of Computer Science and Engineering and Mathematics at Michigan State University. Thanks go to Dr. Joyce Chai, Dr. Sandeep Kulkarni, Dr. Charles Ofria, Dr. Peter Magyar, Dr. Sakti Pramanik, Dr. William Punch, and Dr. Kurt Stirewalt for wonderful experiences in their classrooms. Further thanks go to my fellow graduate students for friendship and inspiration: James Daly, Adam Jensen, Farhana Khan, Eric Norige, Courtland VanDam, and others too numerous to mention. Immense gratitude is due to the Computer Science and Engineering support staff, especially Linda Moore and Norma Teague. They were incredibly knowledgeable and patient, which made the Ph.D. process much smoother than it otherwise would have been.

I would also like to thank my friends and loved ones for their support and understanding. Victoria Ackroyd helped to proofread this manuscript, and so many other things. Most of all I want to thank my maternal grandparents, John and Barbara Schoolenberg, for buying me my first computer so long ago, which began my exploration of computing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Applications of graph theory to real world networks often focus on the use of particular graph structures. Maximization or minimization of a given invariant may correlate with increased network reliability. Only considering graphs of a certain class may allow the use of otherwise unavailable polynomial time complexity algorithms. In this chapter we briefly cover basic graph theory terminology, formally define distance-preserving graphs, discuss our motivation for investigating them, and provide an overview of the rest of the dissertation.

1.1 Definitions and Terminology

Unless otherwise stated, we will use Bondy and Murty [7] as our basic reference for graph theoretic terms. Most other introductory graph theory texts should serve the reader just as well, although notation may vary considerably. A complete list of symbols used in this dissertation may be found in Appendix A. Drawings of special graphs under discussion may be found in Appendix B. For sake of brevity we will refer to distance-preserving graphs as *dp*, and distance-hereditary graphs as *dh*.

A *graph* $G = (V, E)$, alternately $G = (V(G), E(G))$, consists of a set V of *vertices*, and a set E of *edges*, plus an *incidence function* ψ_G assigning each edge to exactly two vertices. The number of elements in the vertex and edge sets, $|V|$ and $|E|$, are referred to as the *order* and *size* of the graph, respectively. The graph of order 1 and size 0 is called a *trivial* graph. A pair of vertices are said to be *adjacent* if they are connected by an edge. The *degree* of a vertex is the number of times it is used as an end-vertex of edges in G. The minimum vertex degree in G is denoted $\delta(G)$, and the maximum vertex degree in G is denoted $\Delta(G)$. A vertex of degree 1 is known as a *pendant vertex*, or a *leaf*. Unless stated otherwise, we will consider *simple* graphs only, without any *loops* (edges that connect a vertex to itself) or *parallel edges* (pairs of vertices connected by multiple edges). When discussing networks, the terms node and link are used instead of vertex and edge, respectively.

In some cases, we want to orient the edges in a graph. A *directed graph* $D = (V, A)$, alternately $D = (V(D), A(D))$, consists of a set V of vertices, and a set D of *arcs*, together with an incidence function ψ_D that assigns each arc to an ordered pair of vertices. If D has an arc $\psi_D(a) = (u, v)$, we say that there is a directed edge from u to v. A directed graph is often referred to as a *digraph*.

We need to define some additional terms before we can begin discussing dh and dp graphs. A graph H is a *subgraph* of G, denoted $H \subseteq G$, if $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$, and $\psi_H \subseteq \psi_G$ is restricted to $E(H)$. A (vertex) *induced subgraph* $H \subseteq G$ is one where $E(H)$ contains every possible edge from $E(G)$, based on the choice of $V(H)$. Given a vertex set $X \subseteq V(G)$, we use $G[X]$ to denote the subgraph induced by X. The *complement* of G, denoted $\overline{G}$, is the graph with the vertex set $V(G)$ and the edge set consisting of exactly those edges not present in $E(G)$. We also need to define a number of basic graph classes. A *path* graph P_n is a graph on n vertices whose vertex set can be arranged in a linear sequence such that two vertices share an edge if and only if they are adjacent in the sequence. Similarly, a *cycle* graph C_n is a graph on n vertices whose vertex set can be arranged in a cyclic sequence such that two vertices share an edge if and only if they are adjacent in the sequence. The *length* of a path or cycle is the number of edges it has. The length of a shortest path between vertices u and v is denoted by $d_G(u,v)$. A *cut vertex* is a vertex whose removal disconnects the graph. A graph is *connected* if every pair of vertices are joined by a path, and *acyclic* if it has no cycles. A graph that is connected and acyclic is a *tree*. The *star* graph S_n is a tree with one vertex connected to $n-1$ other vertices.

A nontrivial graph is *k-connected* if there are k internally disjoint paths between any pair of vertices u and v. In the *complete* graph on n vertices, denoted K_n, every pair of vertices shares an edge. A *bipartite* graph is a graph whose vertices can be partitioned into two disjoint sets U and V such that every edge connects a vertex in U to a vertex in V. If every vertex in U shares an edge with every vertex in V, then the graph is a *complete bipartite* graph. The complete bipartite graph with partitions of order m and n is denoted by $K_{m,n}$.

We might wish to weight or color the vertex and edge sets in a graph, or use directed arcs instead of undirected edges. These variations will be discussed later as needed. However, we now have enough terminology to define the class of dp graphs.

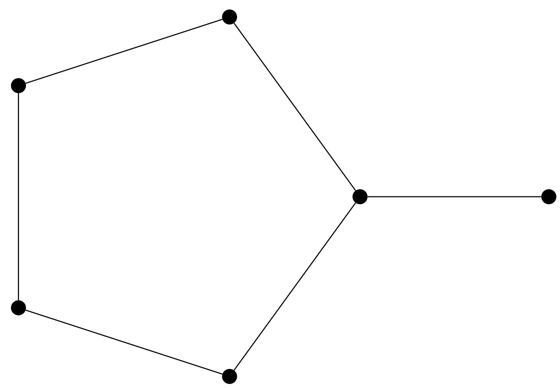
Definition 1.1. *Let G be a graph on n vertices, and H a subgraph of G. We say that H is a dp (isometric) subgraph of G if $d_H(u,v) = d_G(u,v)$ for every $u, v \in H$.*

A *dh* graph is one in which the distances in every connected induced subgraph are the same as they are in the original graph [43], i.e., every connected induced subgraph is isometric. Dh graphs have been studied extensively in the literature; a brief summary is given in Chapter 6. It suffices for now to say that they are a much "stronger" version of our dp graphs. The lowest order graph that is dp but not dh is shown in Figure 1.1.

Definition 1.2. *Let G be a graph on n vertices. We say that G is a dp graph if for each integer k, $1 \leq k \leq n$, there exists a k-vertex isometric subgraph.*

When working with dp graphs and subgraphs care must be taken with terminology. Consider a graph G and subgraph $H \subseteq G$. G may or may not be a dp graph. In either case, H may or may not be a dp subgraph. Furthermore, H may or may not be a dp

Figure 1.1 A Non-DH DP Graph (5-Pan)



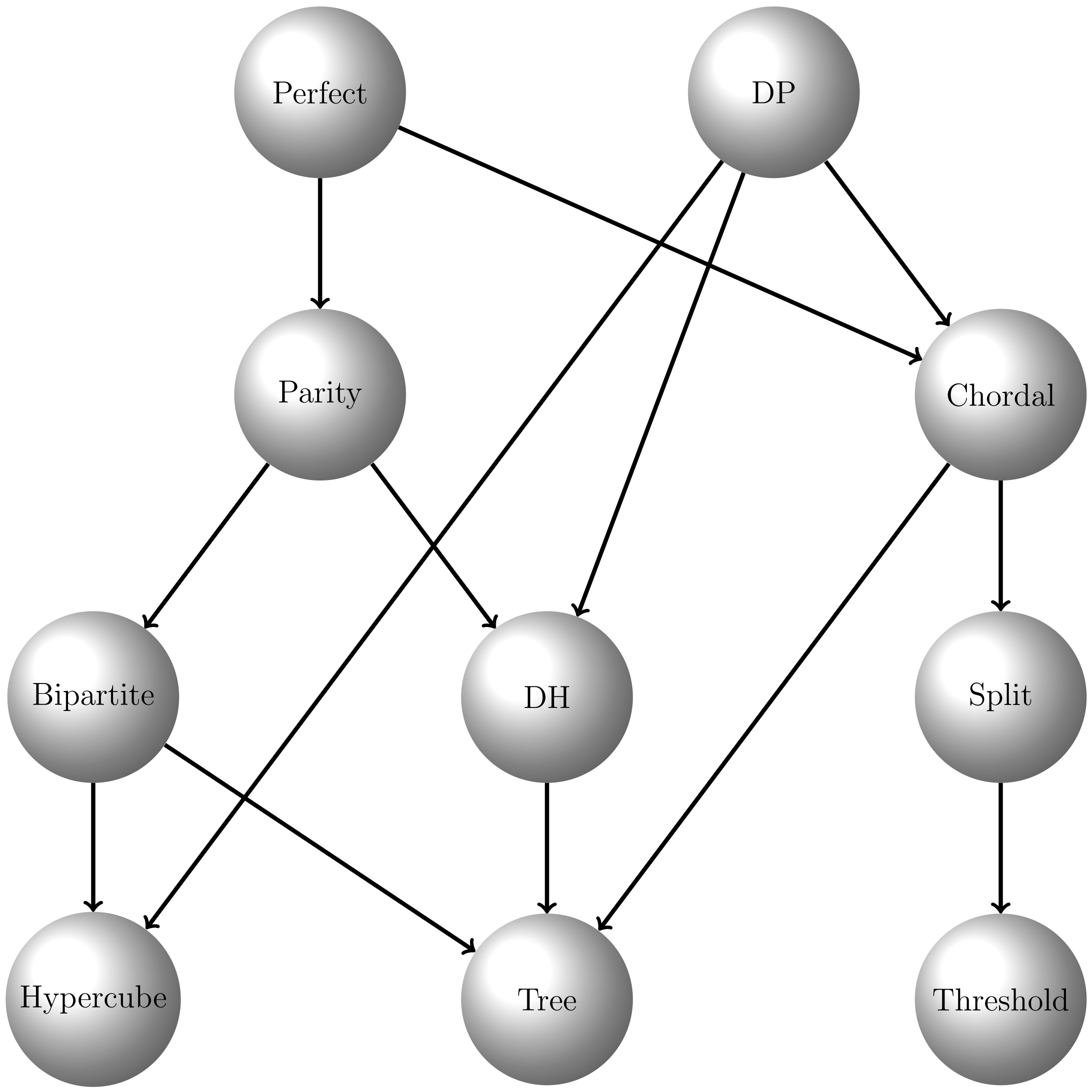
graph in its own right. To avoid any confusion, we will henceforth refer to dp subgraphs as isometric subgraphs.

1.2 Goals and Motivations

Proposing a new class of graphs provides us with some obvious and immediate goals. Alternate characterizations give a deeper understanding of dp graphs. Beyond the connection to dh graphs, we want to know how dp graphs are related to other graph classes. Figure 1.2 [8, 58, 70] shows the known relationships between dp graphs and other known graph classes. Algorithms for recognizing dp graphs and other problems that are NP-Complete for the case of all graphs, or proofs that these are NP-Complete, would be useful as well. Our initial results indicate that recognition and other problems are likely to be NP-Complete. If this is the case, we will want to develop heuristics for finding isometric subgraphs of arbitrary order within a graph.

Regardless of the difficulty, finding isometric subgraphs is an important goal, and perhaps the most likely to lead to practical applications. However, we are also interested in extremal cases, such as the problem of determining the minimum number of additional edges required to turn a non-dp graph into a dp one; a trivial upper bound for this is

Figure 1.2 Graph Classes Hierarchy



$|V(G)| - \delta(G) - 1$. Ultimately, these pursuits may prove more fruitful when working with actual networks.

Another significant goal is finding construction methods for distance-preserving graphs. We know a number of constraints for which this can be done, with varying degrees of difficulty. We have a construction algorithm that works for certain arbitrary degree sequences.

1.3 Isometry in Real-World Networks

Our hypothesis is that isometric subgraphs have applications to real world networks. A low order connected subgraph is very likely to be isometric. A high order isometric subgraph may not be particularly useful in all cases, such as when a graph has few or no cycles. But our expectation is that dense isometric subgraphs of larger order will provide insight into a graph. For example, with the graph of a social network, we expect that the centers of social groups would be dense, with fewer edges between different communities. If this is the case, then partitioning the graph into disjoint isometric subgraphs could provide the basis for a community finding algorithm.

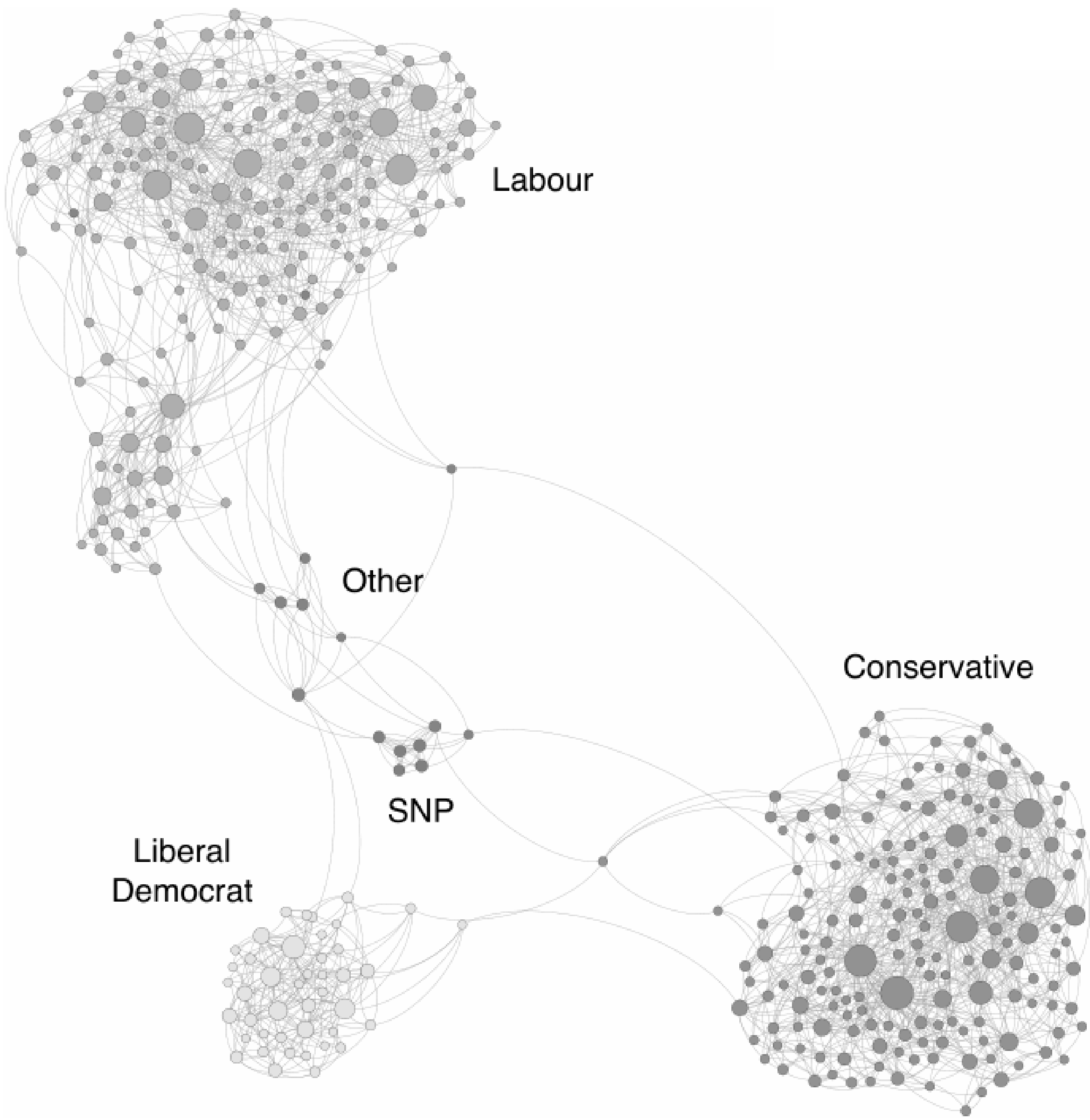
Definition 1.3. *We define the average distance increase for a subgraph of G as the sum of the distance increases between the subgraph and G, divided by the number of vertices in the subgraph. If the subgraph is distance-preserving, then this value is 0. The less distance-preserving the subgraph is, the higher this number will be.*

To help validate our hypothesis we examined five small social media datasets from `http://mlg.ucd.ie/networks/index.html`. Each dataset is a subset of Twitter users representing a network of similar communities. Two are political networks with party affiliations, and three are sports networks with sport / team affiliations:

- United Kingdom Members of Parliament
- Irish politicians and organizations
- Premier League players and clubs
- Olympic atheletes and organizations
- Rugby players and clubs

Each node in these datasets represents a Twitter account. Nodes u, v share a link if u is following v, v is following u, or both. Nodes are assigned to one of several disjoint communities based on political party or team affiliation. The node lists and ground-truth communities were curated manually. Figure 1.3 [37] provides a visualization of one of the datasets.

Figure 1.3 United Kingdom Members of Parliament on Twitter



For each dataset we considered the subgraphs induced by the vertex sets for each class label. The data for these tests may be found in Tables 1-5 in Appendix C. In all cases these subgraphs were isometric, or very nearly so. However, there were a few instances where a subgraph had a few disconnected nodes, such as the Labour party in the UK.

Since these networks were fairly small, and the communities for political parties and sports teams might have fewer edges between them than other sorts of communities, we performed the same tests on two slightly larger networks from `http://linqs.cs.umd.edu/projects//projects/lbc/`. CiteSeer and Cora are paper citation networks. Each node represents a publication. Nodes u, v share a link if u cites v, or v cites u. Nodes are assigned to one of several disjoint communities based on topic. The data for these tests may be found in Tables 6 and 7 in Appendix C. For these datasets, some of the class labels induced clusters that were very nearly distance-preserving with few infinite paths. A few clusters saw distance increases, and some of them had large numbers of disconnected vertices.

1.4 Overview

In this chapter we introduced the notion of dp graphs and isometric subgraphs, addressed our motivations for studying them, and provided some definitions and background. In Chapter 2 we give a formal problem statement for dp graphs, along with some other important questions about them. In Chapter 3 we discuss our observations and characterizations of dp graphs. The role of the four subgraphs forbidden to dh graphs in dp graphs is considered, along with the role that various types of vertices play. In Chapter 4 we focus on constructing dp and non-dp graphs given certain invariants. We show how to construct regular dp graphs, as well as dp graphs for a subset of arbitrary degree sequences. We also address augmenting dp graphs. In Chapter 5 we investigate finding isometric subgraphs, and their applications to community finding using the datasets CiteSeer and Cora. In Chapter 6 we discuss related work, specifically perfect graphs, dh graphs, geodetically connected (gc) graphs, and other graph classes. In Chapter 7 we state our conclusions thus far.

Chapter 2

Problem Statement

It immediately follows from the definitions of dh graphs and dp graphs that the former is a subset of the latter. However, the mere presence of one or even many of the four induced subgraphs forbidden to dh graphs - the house, gem, domino, and long cycle - does not mean that a given graph is not dp. So we are unable to claim that a graph is not dp by looking locally. Furthermore, we know that each non-dp graph must contain at least one of the forbidden subgraphs. For more information about dh graphs and the four aforementioned forbidden subgraphs, see Section 6.2 and Figure 3.1.

Let G be a graph on n vertices, and k an integer, where $1 \leq k \leq n$. Our primary concern is answering the following question. Is there an isometric subgraph $H \subseteq G$ of order k? If we can come up with an algorithm to determine whether G contains an isometric subgraph of arbitrary order, we can use $O(n)$ invocations of the algorithm to determine whether G is dp. If it is possible to do this in polynomial time this would greatly expand the potential applications for dp graphs. In the event that this problem is NP-Complete, our interest would focus on specific cases which have a polynomial time complexity.

We have a number of other questions regarding characterization. Foremost, any classes of graphs that are non-trivially dp. We also want to know more about the roles certain types of vertices play in determining whether a graph is dp, e.g., vertices whose deletion from a graph does not change any other distances. Similarly, there might be operations we can perform on graphs to make it easier identifying whether the original graph is dp or not.

The construction of dp graphs is another concern. Let n be the desired number of vertices and C be some constraint on the graph, such as a specific degree sequence. Can a dp graph G on n vertices be constructed which satisfies condition C? Answering this question for a wide variety of conditions will not only provide more insight into dp graphs, but such constructions may have applications as well. Ultimately, we would like to enumerate all dp graphs give some condition C. The construction of non-dp graphs is another variation on this question.

Another construction consideration involves the addition of edges to a non-dp graph such that the resulting graph is dp. Let G be a non-dp graph. What is a smallest set of edges E such that $G' = G + E$ is a dp graph? E must always exist since G can always be transformed into having a vertex with degree $|V(G)| - 1$. Does each non-complete dp graph have at least one edge that can be added such that the resulting graph is dp? If not, what is the bound on the size of E? Variations on this problem, such as the maximum number of edges that can be added to a non-dp graph such that the resulting graph is still non-dp, might also provide insight into dp graphs.

One last major concern involves the nature of isometric subgraphs in a dp graph. While our first priority is simply to determine whether a graph has an isometric subgraph of a given order order, we would also like to know the bound on the number of isometric subgraphs the graph must contain, and any other characteristics they may possess. Let n be the desired number of vertices and C be some constraint on graph. What are the upper and lower bounds for the number of isometric subgraphs of order k, where $1 \leq k \leq n$? This problem yields two important questions. First, do all graphs have isometric subgraphs for values of k besides 1, 2, 3, 4, and n? For an explanation of why these are trivial, see Section 6.2. Second, do all dp graphs have a large number of isometric subgraphs for $3 \leq k \leq n-2$, say at least $2k$?

We have made considerable amounts of progress in addressing these questions. While we do not yet know the time complexity of the recognition problem, we have a number of sufficient conditions for a graph to be dp. Some of these are bounds on invariants, while others are well known classes of graphs. Thus far our investigation into adding edges to dp graphs has been largely experimental. However, we have shown that when G is sequentially dp, there does always exist an edge e such that $G + e$ is dp. Along with Ross et al., we have developed a construction algorithm to generate regular dp graphs of arbitrary order and vertex degree for which it is possible to construct a dp graph. We have made further progress into generating dp graphs given an arbitrary degree sequence. As with the edge augmentation problem, our investigation into bounds on the number of isometric subgraphs has been largely experimental thus far.

Chapter 3

Characterizations

In our investigation of dp graphs we seek to identify conditions which are sufficient to ensure that a graph is dp. Moreover, we want to determine the exact role the induced subgraphs forbidden to dh graphs play in dp graphs. Our observations indicate that the four forbidden subgraphs are not equally "bad" when it comes to dp graphs. We provide a number of sufficient conditions for a graph to be dp in Section 3.1, and conjecture a number of other conditions that we believe to be sufficient in Sections 3.1 and 3.2. Another avenue of investigation is the existence of special vertices in dp graphs, and the complexity of the recognition problem for dp graphs.

3.1 Observations

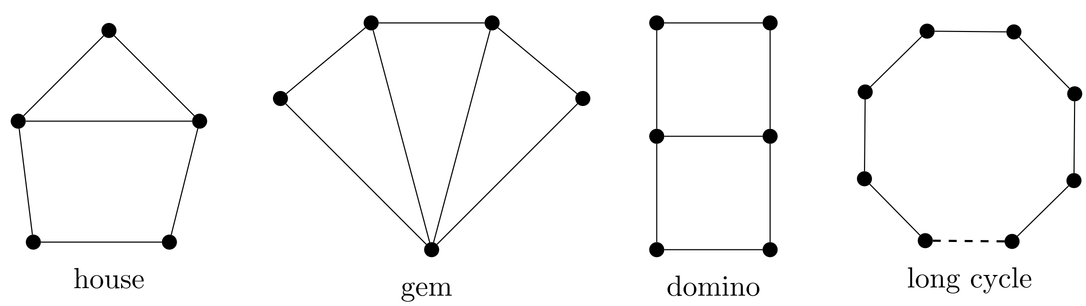
We know that all subclasses of dh graphs are dp graphs. The best known of these are trees, but there are many other such classes [8]:

- cograph ([20])
- ptolemaic ([44, 49])
- threshold ([17])
- trivially perfect ([34])

3.1.1 Forbidden Subgraphs

We first examine the four forbidden subgraphs, shown in Figure 3.1. Whenever we are talking about these graphs as subgraphs of a graph rather than as stand-alone graphs, we are always referring to induced subgraphs. Of these, the long cycle is the only one that is not itself a dp graph. Specifically, C_5 lacks an isometric subgraph of order 4. More generally, we observe that the cycle graph C_n lacks isometric subgraphs of order $\lceil \frac{n}{2} \rceil + 1$ up to order $n - 1$. We also note that the existence of a single connected induced subgraph which is non-isometric will ensure that a graph is non-dh, so all four forbidden subgraphs are on equal ground in this respect. However, the presence of a gem creates 1 non-isometric connected induced subgraph in a graph, while a house or a gem create 2 of these, and a long cycle creates at least 5.

Figure 3.1 Forbidden Subgraphs



Adding just one crossing chord to the 5-cycle allows for an isometric subgraph of order 4, and so the house and the gem are dp. Likewise, the domino is also dp. So the presence of long induced cycles in a graph would appear to be somehow more "disruptive" than the other three forbidden subgraphs. In fact, Conjecture 3.1 is that the presence of a long induced cycle is a necessary condition for a graph to be non-dp.

Conjecture 3.1. *Let G be a graph. If G does not contain an induced subgraph H, where H is a cycle of length 5 or greater (a long induced cycle), then G is dp.*

The conjecture that every non-dp graph must contain an long induced cycle has been difficult to prove so far. Intuitively, we felt that when attempting to add each vertex, we should be able to separate the problem into a few cases, and give a simple proof for each. This might be a direct proof, or a proof by contradiction which relied on the fact that the graph didn't contain any long cycles. After a considerable amount of effort, no counterexample has been found. We have computationally verified that the following classes of graphs are not counterexamples to the Conjecture 3.1:

- all graphs of order $n \leq 12$
- all regular graphs of order $n \leq 13$
- C_3 and C_4 free graphs of order $n \leq 13$
- bipartite graphs of order $n \leq 13$

Randomly searching through graphs of order 12+ has so far been futile.

Of course, a graph can be dp in the presence of an long induced cycle, or even many long induced cycles. In fact, we can construct a dp graph that contains an arbitrary number of induced subgraphs, forbidden or otherwise, as we will see later in Lemma 3.10.

Table 3.1 Percentage of DH Graphs

n	# dh graphs	# connected unlabeled graphs	% dh graphs
1	1	1	100
2	1	1	100
3	2	2	100
4	6	6	100
5	18	21	85.7143
6	73	112	65.1786
7	308	853	36.1079
8	1484	11117	13.3489
9	7492	261080	2.8696
10	40010	11716571	0.3415

Another difference between dh and dp graphs is the percentage of all graphs that are dh and dp. Almost all graphs contain an arbitrary induced subgraph [25], and we see in Table 3.1 how quickly the percentage of graphs which are dh approaches 0 as the number of vertices increase. The fact that dp graphs are so much more common leads us to make Conjecture 3.2.

Conjecture 3.2. *Almost all graphs are dp. That is, as n goes to infinity, the percentage of graphs G of order n that are dp converges to* 1.

Experimentally, the results in Table 3.2 would indicate that this converges fairly quickly, but we would like to prove even some weak lower / upper bounds on n.

Table 3.2 Percentage of DP Graphs

n	# dp graphs	# connected unlabeled graphs	% dp graphs
1	1	1	100
2	1	1	100
3	2	2	100
4	6	6	100
5	20	21	95.2381
6	111	112	99.1071
7	849	853	99.5311
8	11098	11117	99.8291
9	260897	261080	99.9299
10	11714097	11716571	99.9789

The Erdős-Rényi random graph model $G(n, p)$ [29] constructs a graph G on n vertices by connecting each pair of vertices in the graph with probability p. We now ask if the percentage of random graphs that are dp approaches 1 even for arbitrary edge probabilities. In Table 3.3 we compute the percentage of such randomly generated graphs that were dp at regular probability intervals. We do the same for the percolation theshold ($p = \frac{1}{n}$) [38] and the connectivity threshold ($p = \frac{\log n}{n}$)[29], running 100,000 trials for each combination of n and p. For all values of p that we examined, almost all graphs are dp.

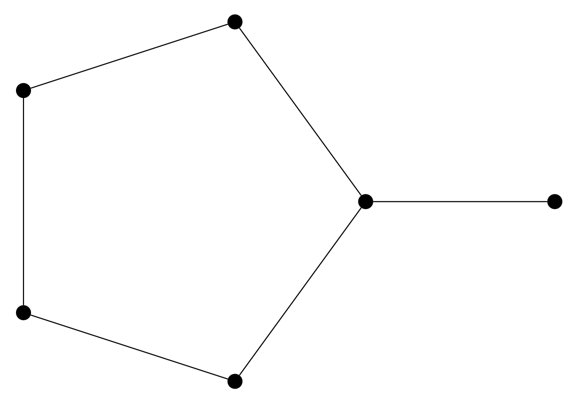
Table 3.3 Percentage of Erdős-Rényi Graphs Which Are DP

$n \backslash p$	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
5	1	1	1	.999	.997	.995	.995	.990	.988	.987
6	1	1	1	1	.999	.998	.998	.998	.998	.998
7	1	1	1	1	.999	. 998	.998	.998	.998	.999
8	1	1	1	.999	.997	.997	.997	.997	.998	.999

3.1.2 Sequentially DP Graphs

In a *sequentially* dp graph, each subgraph in the set of isometric subgraphs is a superset of the previous one. In other words, there exists some permutation of the vertices such that the subgraphs induced by the first $1, 2, \ldots$ vertices in the permutation are all isometric. A sequentially dp graph is necessarily dp, but the converse is not always true. The lowest order example of dp graph which is not sequentially dp is the $5 - Pan$ graph shown in Figure 3.2, which consists of C_5 plus an additional pendant vertex.

Figure 3.2 A Non-Sequential DP Graph (5-Pan)



3.1.3 Special Vertices

From the definition of a dp graph we know that it must have at least one vertex which may be deleted from the graph without increasing any distances. A vertex in a graph may be contained in isometric subgraphs of every possible order, or significantly fewer isometric subgraphs than that. We consider vertices which exist in isometric subgraphs of every possible order, and vertices for which a set of isometric subgraphs of every possible order exist without that vertex. We want to determine how many of these vertices a graph must contain to be dp, if any, and their nature. Knowing more about these kinds of vertices could help prove conjectures, and to characterize graph operations on dp graphs.

Definition 3.3. *Let G be a graph on n vertices, and v a vertex in G. We call v an anchor vertex if G contains a k-order isometric containing v for $1 \leq k \leq n$.*

Table 3.4 Percentage of Vertices in DP Graphs of Order n Which Are Anchor Vertices

n	% anchor
3	100
4	100
5	100
6	99.6997
7	99.9159
8	99.9212

From Definition 3.3, a graph with an anchor vertex must be a dp graph. We see in Table 3.4 that the majority of vertices in dp graphs are anchor vertices. We want to determine if the converse is true. In Table 3.5 we count the number of anchor vertices for every graph of a given order, for $n \leq 8$.

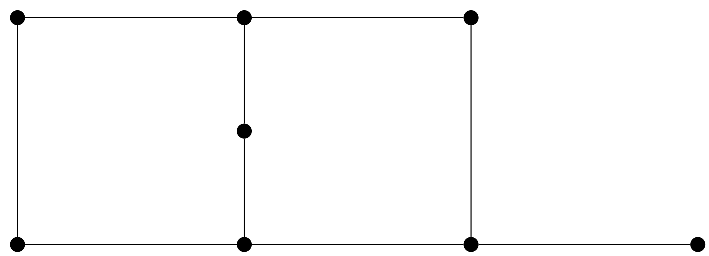
Table 3.5 Number of DP Graphs of Order n With Exactly k Anchor Vertices

$n \backslash k$	1	2	3	4	5	6	7	8
3	0	0	2					
4	0	0	0	6				
5	0	0	0	0	5			
6	0	0	0	0	2	109		
7	0	0	0	0	0	5	844	
8	0	0	0	0	3	4	54	11037

In fact, every dp graph of order $5 \leq n \leq 8$ contains at least 5 anchor vertices. So does the subset of dp graphs of orders $9 \leq n \leq 12$ we checked, as well as higher order graphs from the House of Graphs dataset [9] at http://hog.grinvin.org. Since so many of the vertices in dp graphs are anchor vertices, it is hard to say much about them, although it allows us to make Conjecture 3.4.

Conjecture 3.4. *Every dp graph contains at least one anchor vertex.*

Figure 3.3 A DP Graph With Central Anchor Vertices



Non-anchor vertices often belong to the periphery of a graph, but not always. In the small order graph of Figure 3.3, the subgraph induced by the set of anchor vertices was always connected.

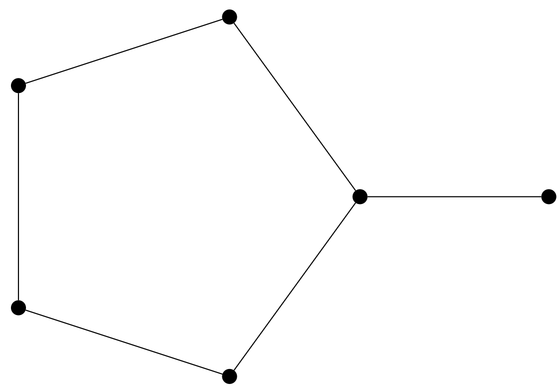
Definition 3.5. *Let G be a graph on n vertices, and v a vertex in G. We call v a nonessential vertex if g contains a k-order isometric subgraph not containing v for $1 < k < n$.*

A nonessential vertex is different from a vertex where the subgraph induced by the remaining vertices is isometric, e.g. a pendant vertex. While a nonessential vertex is by definition the latter, the converse is not true. In Figure 3.4 we again see the 5-pan graph. The pendant vertex is contained in all isometric subgraphs of order 4, and all other vertices comprise the single isometric subgraph of order 5.

Definition 3.6. *Let G be a graph on n vertices, and v a vertex in G. We call v a simplicial vertex if the neighbors of v are a clique and $G - v$ is an isometric subgraph of G.*

We are, however, interested in those vertices which can be deleted without increasing any distances in a graph. We call these vertices *removable*, or *non-hinge* vertices. Simplicial vertices, which include pendant vertices, are a subset of these. From the definition we know that each dp graph must have at least one removable vertex. Unfortunately, almost all

Figure 3.4   A DP Graph With No Nonessential Vertices (5-Pan)



graphs do not have a simplicial vertex, so we are forced to find the other removable vertices to show that a graph is dp or sequentially dp. In Tables 3.6 and 3.7, we examine the result of deleting simplicial vertices from dp graphs. We observe that deleting just one simplicial vertex in a low order graph often causes the resulting graph to be non-dp. Our interest in simplicial and similar types of vertices is to develop strategies to solve problems such as Conjecture 3.1.

Table 3.6   Graphs With Simplicial Vertices Whose Deletion Results in a DP Graph

n	# connected graphs with a simplicial vertex whose deletion leaves a dp graph	# connected graphs	% connected graphs with a simplicial vertex whose deletion leaves a dp graph
2	1	1	100.000
3	2	2	100.000
4	5	6	83.333
5	17	21	80.952
6	86	112	76.786
7	660	853	77.374
8	8000	11117	71.962
9	165726	261080	63.477

Table 3.7 Highest Degree Simplicial Vertices in Connected Graphs Whose Deletion Results in a DP Graph

$n\backslash\#$	1	2	3	4	5	6	7	8
2	1	0	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0
4	2	2	1	0	0	0	0	0
5	4	9	3	1	0	0	0	0
6	20	41	20	4	1	0	0	0
7	132	314	171	37	5	1	0	0
8	1478	3653	2312	490	60	6	1	0
9	28064	73582	51412	11390	1179	91	7	1

3.1.4 Hinge-Free Graphs

Definition 3.7. *Let G be a graph on n vertices. We say that G is a k-geodetically-connected (k-gc) graph if every induced $n-k$ order subgraph of G is isometric.*

A *hinge-free* or *1-geodetically connected* (1-gc) graph is a graph composed entirely of removable vertices. For more background on hinge-free graphs, see Section 6.3. Hinge-free graphs necessarily have redundant shortest paths, which leads us to give Conjecture 3.8.

Conjecture 3.8. *Let G be a k-gc graph, where $1 < k < n$. Then G is a dp graph.*

At present we are focused solely on 1-gc graphs, since graphs with $\delta(G) > n/2$ are 1-gc graphs, and a proof for Conjecture 3.8 would also prove Conjecture 3.12. Table 3.8 shows invariant bounds for hinge-free graphs. The extremal cases for most bounds on graph invariants that we looked at are K_n and $K_{n-2,2}$.

Table 3.8 Invariant Bounds for 1-GC Graphs

n	#	$size(G) \geq$	$diam(G) \leq$	$girth(G) \leq$	largest induced cycle
3	1	3	1	3	3
4	3	4	2	4	4
5	7	6	2	4	4
6	30	8	3	4	5
7	141	10	3	4	5
8	1259	12	4	4	6
9	21176	14	4	4	6

As with long induced cycle free graphs, Conjecture 3.9 states that for any vertex in a hinge-free graph, we can find an isometric subgraph of arbitrary order containing that vertex, and another one lacking it. Table 3.9 shows that the minimum number of k-order isometric subgraphs in a graph of order n is quite large.

Conjecture 3.9. *Let G be a hinge-free graph on n vertices, and v a vertex in G. Then G contains k-order isometric subgraphs H and H', such that $v \in H$ and $v \notin H'$, for $1 \leq k \leq n$.*

Table 3.9 Minimum Number of k-order Isometric Subgraphs in Hinge-Free Graphs

$n \backslash k$	1	2	3	4	5	6	7	8	9
3	3	3	1						
4	4	4	4	1					
5	5	6	9	5	1				
6	6	8	14	9	6	1			
7	7	10	21	16	12	7	1		
8	8	12	24	26	24	16	8	1	
9	9	14	33	45	45	32	18	9	1

We might prove the Conjecture 3.8 by showing that every k-order subgraph that maximizes or minimizes some invariant must be isometric. Experimentation has shown that maximizing size or minimizing diameter will not work.

We might also start with a random vertex, and by maximizing or minimizing some invariant as we go, build a set of distance preserving subgraphs incrementally. Adding vertices in the 1-neighborhood, then the 2-neighborhood, and so on, will not work. We would expect maximizing size or minimizing diameter not to work, but we are doing something slightly different than we did before.

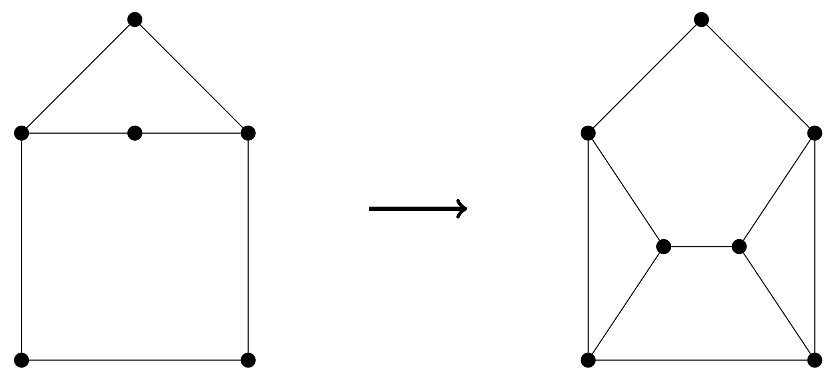
3.1.5 Line Graphs of DP Graphs

Let G be a graph. The line graph [69, 42] of G, denoted by $L(G)$, is defined as follows:

- $|V(L(G))| = |E(G)|$, where each vertex in $L(G)$ corresponding to an edge in $V(G)$.
- Vertices $u, v \in L(G)$ are adjacent if and only if $uv \in E(G)$.

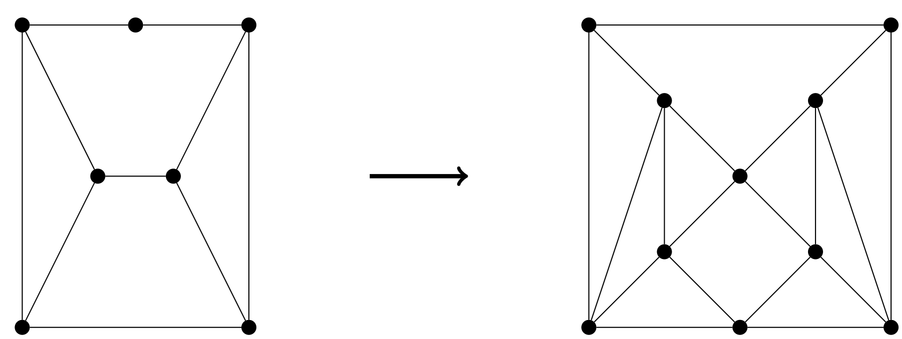
Line graphs may be characterized in a number of ways. Beineke characterized line graphs in terms of forbidden subgraphs, proving that there are only nine minimal graphs that are not line graphs [2, 3]. Furthermore, there is a one-to-one correspondence between isomorphisms of graphs and the isomorphisms of line graphs, for all graphs with more than 4 vertices [24]. This led us to consider the possibility that the line graph of a dp graph might always be dp. Some experimentation quickly shows that this is not the case. Also, the line graph of a non-dp graph is not necessarily non-dp, as seen in Figure 3.5.

Figure 3.5 A DP Graph Whose Line Graph is Non-DP



Similarly, the line graph of a non-dp graph is not necessarily non-dp, as seen in Figure 3.6.

Figure 3.6 A Non-DP Graph Whose Line Graph is DP



3.2 Results

Our first question is on the relationship between dp graphs and other graph classes. All graphs of up to order 4 are both dp and perfect. The only non-dp graph on 5 vertices is C_5, which is also the only non-perfect graph on 5 vertices. However, C_6 is perfect, but not dp. Likewise, each of the graphs containing induced 5-cycles on 6 vertices are dp, but not perfect. So, neither dp graphs nor perfect graphs are subsets of each other.

A simple result highlights the different between dh and dp graphs. Dp graphs cannot be defined strictly in terms of forbidden subgraphs, as shown in Lemma 3.10.

Lemma 3.10. *Given a multiset* $\mathcal{F} = \{F_1, F_2, \ldots, F_p\}$ *of simple graphs of order* $n_1, n_2, \ldots, n_p$, *we can construct a dp graph* G *containing every element of* $\mathcal{F}$ *as an induced subgraph.*

Proof. Let $n = 2 \cdot (n_1 + n_2 + \ldots + n_p)$, G be the disjoint union of $K_{n/2} \cup F_1 \cup F_2 \cup \ldots \cup F_p$ on n vertices, and u be an arbitrary vertex in $K_{n/2}$. Add edges $\{u, f_i\}$, where f_i is an arbitrary vertex in F_i, $1 \leq i \leq p$. We can find isometric subgraphs in G of order 1 up to order $\frac{n}{2}$ simply by choosing arbitrary vertices from the subgraph $K_{n/2}$. We can find isometric subgraphs in G of order $\frac{n}{2} + 1$ up to order $\frac{n}{2} + n_1$ by choosing all the vertices in G corresponding to F_1, plus u and arbitrary other vertices from the subgraph $K_{n/2}$. We can find isometric subgraphs in G of order $\frac{n}{2} + n_1 + 1$ up to order $\frac{n}{2} + n_1 + n_2$ by choosing all the vertices in G corresponding to F_1 and F_2, plus u and arbitrary other vertices from the subgraph $K_{n/2}$. We can continue with this to find isometric subgraphs in G of all the way up to order n. Thus, Q_n is a dp graph. □

3.2.1 Minimum Degree and Size

In Table 3.10 we see that although almost all graphs may be dp, $\delta(G)$ has to be quite high to ensure that G is dp. For $n = 13$ and $n = 14$, we tested likely non-dp candidates where $\delta(G) \leq \Delta(G) + 1$, hence the asterisk. This was done out of computational necessity.

Table 3.10 Minimum $\delta(G)$ Necessary to Ensure G is a DP Graph

n	$\delta(G)$
3	1
4	1
5	3
6	3
7	3
8	4
9	5
10	5
11	5
12	6
13	7*
14	7*

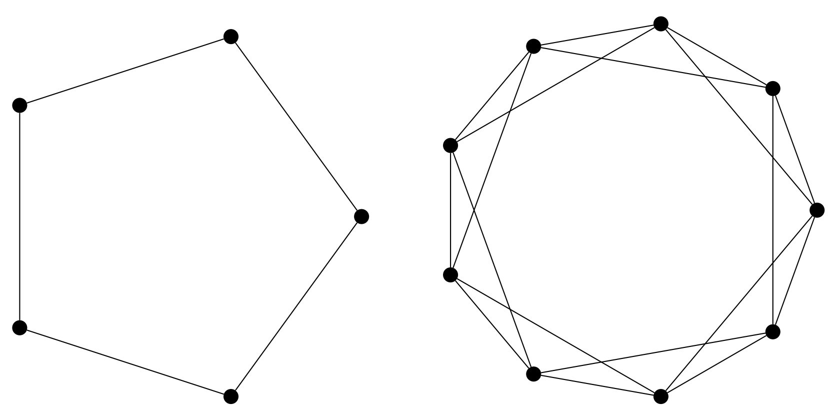
Lemma [58] 3.11. *Let G be an n-vertex graph with $\delta(G) \geq \frac{2n}{3} - 1$. Then G is a dp graph.*

Proof. Let H_k denote the proposed isometric subgraph of order k. G must contain star subgraphs up to at least order $\delta(G) + 1$. For $1 \leq k \leq \frac{2n}{3}$, let H_k be the star subgraph consisting of an arbitrary vertex, and $k - 1$ of its neighbors. Each star subgraph is dp, as the distance between the internal node and each of the leaves is 1 in G and H_k, and the distances between each pair of leaves remains 1 if they share an edge in G, and 2 if they do not. For $\frac{2n}{3} < k \leq n$, let H_k be an induced subgraph containing k arbitrary vertices. Now consider each pair of vertices u, v in these random induced subgraphs. If u and v share an edge in G, $d_G(u, v) = d_{H_k}(u, v) = 1$. If u and v do not share an edge in G, $d_G(u, v) = 2$, since they must have common neighbors. Now consider the the worst case, where $deg(u) = deg(v) = \frac{2n}{3} - 1$. Then there are $n - 2 - (\frac{2n}{3} - 1) = \frac{n}{3} - 1$ vertices u does not share an edge with (besides v). Even if v shares an edges with each of these $\frac{n}{3} - 1$

edges, u and v still have $(\frac{2n}{3} - 1) - (\frac{n}{3} - 1) = \frac{n}{3}$ shared neighbors. Since $k > \frac{2n}{3}$, u and v must have at least one shared neighbor in H_k, which means that $d_{H_k}(u, v) = 2$, and each random induced subgraph where $k > \frac{2n}{3}$ is dp. Thus, G is a dp graph. □

$\delta(G) \geq \lfloor n/2 \rfloor$ is sufficient to ensure that every pair of vertices in G are adjacent or have at least one common neighbor, that G is Hamiltonian [26], and G has diameter of 1 or 2. Every graph with $\delta(G) \geq \lfloor n/2 \rfloor$ contains a triangle, except for $K_{1,1}, K_{2,2}, \ldots$ It is not a sufficient condition for G to be dp if G has odd order ($n = 5, 9, 13, \ldots$), as seen in Figure 3.7.

Figure 3.7 Non-DP Odd Order Graph With $\delta(G) = \lfloor n/2 \rfloor$



Conjecture 3.12. *Let G be an n-vertex graph with $\delta(G) > n/2$. Then G is a dp graph.*

This conjecture is a lower bound on the minimum vertex degree needed to ensure that a graph is dp. Doing this indirectly by specifying the minimum degree of each vertex in the graph might make it easier to prove. Ideally, we would like to go beyond edges and think of bounds in terms of the number and arrangement of forbidden subgraphs, but nothing so far leads one to believe that this is even possible. We have computationally verified that the conjecture holds for $n \leq 12$. We have also ruled out likely counterexamples for $n = 13$ and $n = 14$, which are graphs where $\Delta(G) \leq \lceil n/2 \rceil + 1$. We have further verified that these graphs are sequentially dp for $n \leq 11$ except for the graph in Figure 3.8.

Figure 3.8 Non-DP Odd Order Graphs With $\delta(G) = \lfloor n/2 \rfloor$

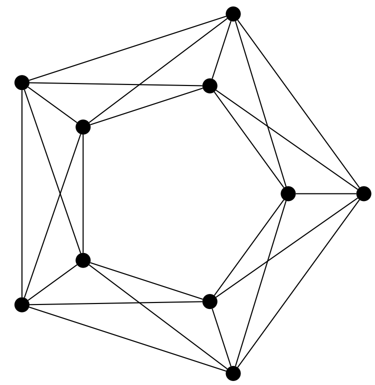
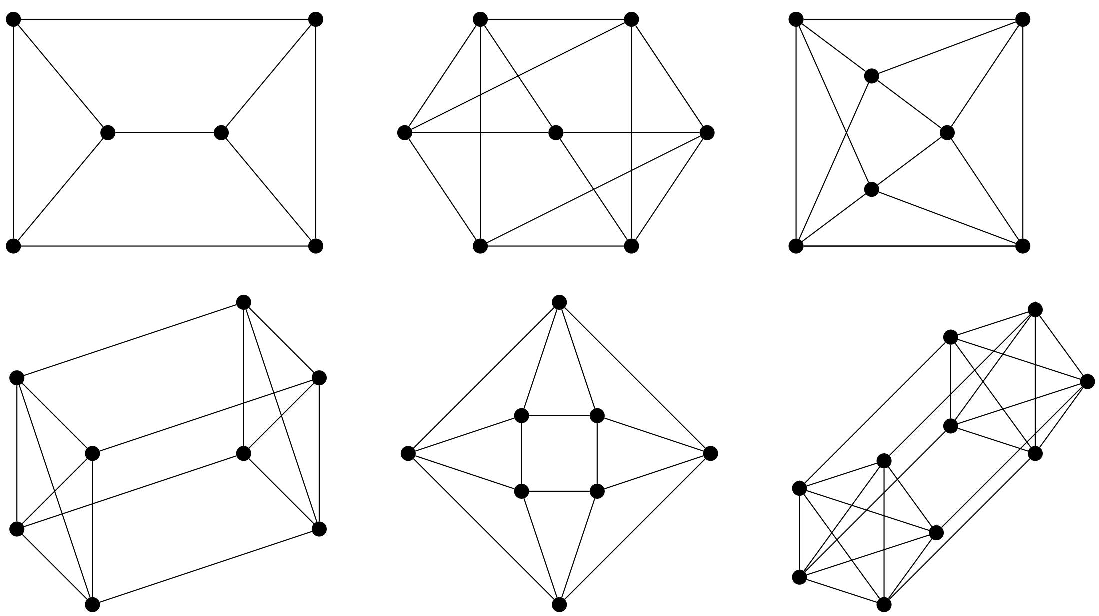


Table 3.11 Minimum Number of Isometric Subgraphs of Graphs of $\delta(G) \geq n/2$

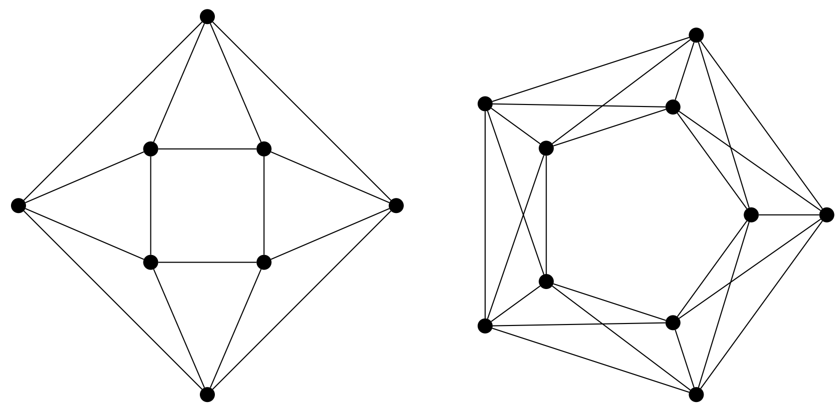
$n \backslash k$	1	2	3	4	5	6	7	8	9	10	11	12
3	3	3	1									
4	4	4	4	1								
5	5	8	10	5	1							
6	6	9	14	9	6	1						
7	7	14	28	28	21	7	1					
8	8	16	32	26	24	16	8	1				
9	9	23	59	76	88	73	36	9	1			
10	10	25	60	65	62	80	60	25	10	1		
11	11	33	99	154	220	252	231	143	55	11	1	
12	12	36	100	159	200	228	244	165	100	36	12	1

Figure 3.9 Graphs With the Minimum Number of Isometric Subgraphs of Order $n-2$



The Cartesian product $K_{n/2} \times K_2$ does not always give a graph with the fewest number of isometric subgraphs for all $1 \leq k \leq n$. For example, $K_4 \times K_2$ contains 32 isometric subgraphs of order 4 and 32 isometric subgraphs of order 5, and $K_5 \times K_2$ contains 80 isometric subgraphs of order 4 and 102 subgraphs of order 5. These graphs have the fewest number of isometric subgraphs of order $n/2$ for $n = 8$ and $n = 10$, respectively ($K_5 \times K_2$ has more than the minimum number of isometric subgraphs for $6 \leq k \leq 8$:

Figure 3.10 Fewest Number of Isometric Subgraphs of Order $n/2$ for $n = 8$ and $n = 10$



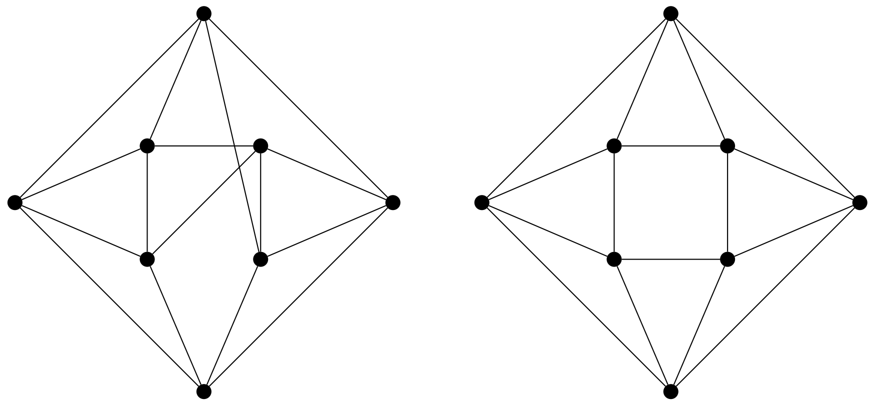
If every graph with $\delta(G) \geq n/2$ contained at least one induced subgraph with minimum degree of at least $\lfloor k/2 \rfloor$, for $1 \leq k \leq n$, we would be done. This is not the case:

Table 3.12 Minimum Maximum $\delta(G)$ of Induced Subgraphs of Graphs of $\delta(G) \geq n/2$

$n \backslash k$	1	2	3	4	5	6	7	8	9	10
3	0	1	2							
4	0	1	1	2						
5	0	1	2	2	3					
6	0	1	1	2	2	3				
7	0	1	2	2	2	3	4			
8	0	1	1	2	2	2	3	4		
9	0	1	2	2	3	3	3	4	5	
10	0	1	1	2	2	3	3	3	4	5

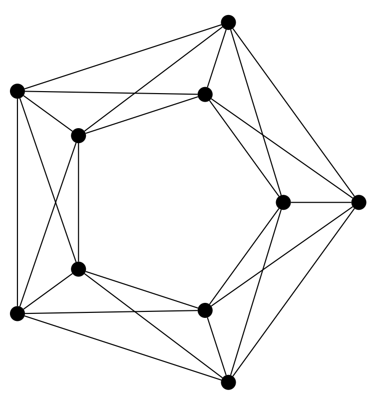
These two graphs of order 8 lack induced subgraphs of order 6 with minimum degree greater than 2:

Figure 3.11 Graphs With the Fewest Number of Isometric Subgraphs of Order $n/2$



In both cases some, but not all, of the maximal minimum degree induced subgraphs of order 6 are isometric. We have not found a graph with $\delta(G) \geq n/2$ where none of the maximal minimum degree induced subgraphs of some order are isometric. Contrast this to subgraph size, where we have found examples of such graphs where none of the maximal size induced subgraphs of some order are isometric. In the graph in Figure 3.12 the maximal size induced subgraphs of order 6 have 13 edges, none of which are isometric:

Figure 3.12 A Graph Where No Maximal Size Induced Subgraph of Some Order is Isometric



Looking for maximal maximum degree induced subgraphs of a given order produces too many counterexamples on 8 and 10 vertices to list.

Conjecture 3.13. *Let G be a simple graph of order n and size m. If $m > n \cdot (n-1)/4$, then G is dp.*

This is a stronger version of Conjecture 3.12. Rather than requiring every individual vertex in G to have over half the number of potential edges, this conjecture only requires that G as a whole contain half the number of potential edges. As previously mentioned, we expect that this conjecture will be harder to prove. Again, this conjecture has been experimentally verified for graphs of up to 10 vertices, as seen in Table 3.13.

Table 3.13 Minimum Size Necessary to Ensure G is a DP Graph

n	size
5	6
6	7
7	11
8	13
9	19
10	22

Conjecture 3.14. *Let G be a graph. Except for C_5, at least one of G or $\overline{G}$ is a dp graph.*

Thus, all self-complementary graphs, except for C_5, must be dp. If true, this conjecture will likely follow as a corollary of a Conjecture 3.13 regarding the number of edges in a dp graph. This conjecture has been verified experimentally for all graphs of up to order 9.

The only non-trivial superclass of dp graphs currently known is the set of all simple graphs.

3.2.2 Hypercubes

By making use of the binary labeling of a hypercube, we show that hypercubes are dp. We suspect there are other methods for selecting a set of isometric subgraphs in hypercubes, and that generalized de Dbruijin graphs on a binary alphabet are dp as well.

Theorem 3.15. *Consider the n-dimensional hypercube Q_n, with standard n-bit binary string vertex labeling. Then the induced subgraph on the first k vertices in lexicographic order, denoted H_k, is an isometric subgraph of Q_n, for $1 \leq k \leq 2^n$. So Q_n is a dp graph.*

Proof. Proof is by induction on k.

Basis: For $k = 1$, H_1 is trivially dp, since it contains only the zero length path starting at v_0.

Inductive Hypothesis: Now assume the statement holds true for $k = 1, 2, \ldots, K$.

Inductive Step: For $K + 1$, we know from applying the inductive hypothesis that $distance_{H_{K+1}}(v_i, v_j) = distance_G(v_i, v_j)$ for all $v_i, v_j \in H_{K+1} \backslash \{v_K\}$. So we only need concern ourselves with the shortest paths in H_{K+1} between v_K and v_i. We know from the definition of the hypercube that $distance_{Q_n}(v_i, v_K)$ is equal to the total number of bits that differ between the labels of v_i and v_K. Now we will create a path of the same length in H_{K+1}, starting with v_K. If there are any bits that are 1 in the label for v_K and 0 in the label for v_i, we can flip these bits from 1 to 0 one at a time in some arbitrary order, adding each corresponding vertex to the path in turn. Every vertex traversed so must be found in H_{K+1}, since they all come lexicographically before v_K. After this step is complete, if there are any bits that are 0 in the label for v_K and 1 in the label for v_i, we can flip these bits from 0 to 1 one at a time in some arbitrary order, again adding each corresponding vertex to the path. Since v_i comes lexicographically before v_K, the position of the leftmost bit we flip from 0 to 1 must be to the right of the leftmost bit that we flipped from 1 to 0, and each of the vertices added in this step must come lexicographically before v_K as well. Because the length of our path is equal to the total number of bits that differ between the labels of v_i and v_K, $distance_{H_{K+1}}(v_K, v_i) = distance_{Q_n}(v_K, v_i)$. So H_{K+1} is an isometric subgraph of Q_n, and our inductive step is complete.

Thus G is a dp graph. □

We note that hypercubes are not only dp, but sequentially dp. If the dp qualities of hypercubes have practical applications, other sequential dp orderings may be of interest; consider Conjecture 3.16.

Conjecture 3.16. *Consider the n-dimensional hypercube Q_n, with standard n-bit binary string vertex labeling. Then the induced subgraph on the first k vertices in reflected binary code (Gray code) order, denoted H_k, is an isometric subgraph of Q_n, for $1 \leq k \leq 2^n$. So Q_n is a dp graph.*

We believe undirected generalized de Bruijn graphs over binary alphabets are dp as well. We have computationally verified that $B(2, n)$ is dp for $1 \leq n \leq 4$.

Conjecture 3.17. *The undirected generalized n-dimensional de Bruijn graph $B_{2,n}$ on the alphabet $\{0, 1\}$ is a dp graph.*

3.3 Complexity

Proving which complexity class the dp graph problem belongs to may be very difficult. As with most problems that belong to NP, it is straightforward to show that the decision problem for dp graphs does in fact belong to NP. The problem of finding the longest cycle in a graph, itself transformed from the Hamiltonian cycle problem, has the most superficial resemblance to the dp graph problem. As a dp graph may have long induced cycles of arbitrary sizes, performing a reduction between the two is not a simple matter, and it is not clear how we might split / add / remove vertices and edges to accomplish a correct reduction. It appears to be difficult or impossible to do this without knowledge of how to determine whether a graph is dp from its long induced cycles.

Lemma 3.18. *The decision problem for dp graphs belongs to NP.*

Input. A simple graph G on n vertices, and k an integer, where $1 < k < n$.

Question. Is there a subgraph $H \subseteq G$ of order k such that H is an isometric subgraph of G?

Proof. We provide an appropriate certificate and verification algorithm.

Certificate. The certificate is a subgraph $H \subseteq G$ of order k.

Algorithm. We use the Floyd-Warshall algorithm (or any other all-pairs shortest paths algorithm requiring polynomial time) on G and H. If any pair of vertices $u, v \in H$ such that $distance_H(u, v) \geq distance_G(u, v)$, then the algorithm returns no. Otherwise, the algorithm returns yes.

Since a solution to the dp graph problem may be verified in polynomial time, the problem belongs to NP. □

Conjecture 3.19. *Deciding whether a graph G contains an isometric subgraph of order k, for $1 \leq k \leq n$, is NP-Complete.*

Assuming the isometric subgraph problem is NP-Complete, we have several choices. We can:

- Look for special cases.
- Relax our requirements.
- Find a worst case EXPTIME algorithm that is fast in practice.

Examination of special cases seems least likely to be useful for practical application, although it may be of some theoretical interest. For some cases, such as planar graphs, generating fundamental cycles can be done in $O(E)$ time. Relaxing the requirements is probably the most suitable option here, and that is what we did in [59]. It seems that any exponential time algorithm should run too slowly to be useful even for very sparse social network datasets, but it is a possibility.

Our first attempts at exploring dp graphs were done through attempting to build up large isometric subgraphs in an incremental fashion. We learned that maximizing or minimizing certain graph invariants led to larger isometric subgraphs before we got "stuck". If our long cycle conjecture holds, it is reasonable to believe that diameter minimization should find good clusters even in non-dp graphs. Unfortunately, doing this in the context of a clustering algorithms requires $O(n^5)$ or $O(n^6)$ time complexity, which is too slow to be useful in practice. We might try to overcome this hurdle by resorting to easier to calculate measures that behave similarly to minimizing the diameter of our cluster.

3.4 Summary

We have found a number of results that help us to characterize dp graphs, which have in turn led to more open problems. Although the most pressing of these now are Conjecture 3.1 and the question of NP-Completeness, we have a number of other important questions. We would like to formally prove Conjecture 3.2, which states that the percentage of graphs that are dp converges to 1 as the number of vertices approaches infinity. Many of our conjectured possible alternate characterizations right now are extremal in nature. We would like to find some radically different characterizations as well, although there are probably not any that only consider forbidden subgraphs.

Chapter 4

Constructions

In this chapter we consider the problem of constructing dp graphs. If we have a dp graph, we would like to know what edges can be added such that the augmented graph is dp. We do not consider the addition of vertices, as a dp graph plus an isolated or pendant vertex will always be dp. Given some constraints, we would like algorithms that generate dp graphs satisfying those constraints, assuming any exist. As with finding characterizations this process provides us with more insight into the nature of dp graphs. We also look at the problem of adding edges to non-dp graphs so that the augmented graph is dp, and constructing non-dp graphs for some given constraints. Unless otherwise stated, we are only considering connected constructions.

4.1 Observations

For certain constraints such as order, size, diameter, radius, etc., finding a construction is trivial. For each of these we can simply take the appropriate order path graph, or another tree. For other constraints it is less obvious what a general algorithm might look like. Combinations of constraints increase the difficulty of finding a construction, although not always by very much. For example, in Lemma 3.11 we used the fact that induced subgraphs of a graph which have an underlying star subgraph are isometric. Similarly, a graph with an underlying star subgraph must be dp. To construct a dp graph of order n and size m, we can start with the star graph S_n and add edges at random until the resulting graph is of size m.

Our primary goal here is to find a construction algorithm that can generate a dp graph given an arbitrary graphical degree sequence for which any dp graphs exist. Table 3.2 gives some credence to the possibility that the construction problem should be straightforward for any constraints for which a deterministic construction method exists. This problem proved quite challenging, so we first look at specific cases.

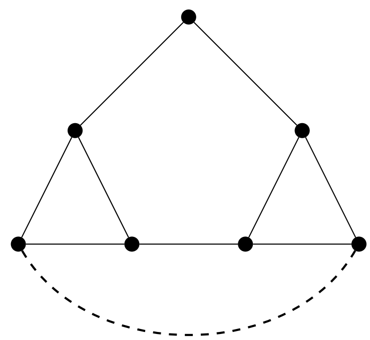
4.2 Altering DH and DP Graphs

In this section we examine adding edges from dh and dp graphs while leaving them dh and dp. We give a number of results for these problems. We also consider the minimum number of additional edges needed to make a non-dp graph into a dp one. So far we have no new conjectures in this regard, although there are some results regarding the former problem for special cases of graphs [31].

4.2.1 Adding Edges to DH and DP Graphs

As with a dh graph, adding an arbitrary edge to a dp graph may make the augmented graph non-dp. For example, consider adding an edge connecting the two leaves in P_5, or the graph in Figure 4.1. However, there exists an edge e that we can add to P_5 such that $P_5 + e$ is dp. We would like to show that such an edge exists for an arbitrary dp graph. Towards this end, we begin with Theorem 4.1.

Figure 4.1 An Edge Whose Addition Makes a DP Graph Non-DP



Theorem 4.1. *Let G be a non-complete dh graph with at least one bridge or a non-complete block. Then there exists some new edge e such that $G + e$ is dh.*

Proof. Let G be a non-complete dh graph with at least one bridge or a non-complete block. We adopt a simple strategy for selecting an edge to add to G.

Case 1: Every block in G is complete. Let $e = (u, v)$ be any edge not in $E(G)$ where $d_G(u, v) = 2$, where uw, vw, or both are bridges. Either w will be in the same block as u, v, or both, or w will itself be part of another block. Regardless, adding e to G will not create any of the four forbidden subgraphs.

Case 2: At least one block in G is not complete. Let $H \subseteq G$ be one such component, and $e = (u, v)$ be any edge where $d_G(u, v) = 2$ and $u, v \in V(H)$. We note that every pair of non-adjacent vertices in the same block in a dh graph with one common neighbor must have two common neighbors. Let two arbitrary common neighbors of u and v be denoted by w and x. We show by contradiction that if $G + e$ were to contain one of the four forbidden subgraphs, G would as well. Suppose that $G + e$ contains one of the following induced forbidden subgraphs:

House: G must contain a domino or at least 2 induced 5-cycles.

Gem: G must contain at least one house.

Domino: G must contain at least two induced 5-cycles, and possibly an induced 6-cycle.

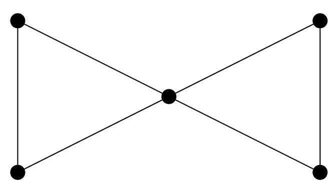
Long Cycle: G must contain at least one gem, house, domino, or long induced cycle.

In each of the four cases we arrive at a contradiction. Once again, $G + e$ must not contain any of the four forbidden subgraphs.

Thus $G + e$ is a dh graph. □

We note that if a dh graph G contains no bridge edges and all blocks are complete, there are no edges e such that $G + e$ is dp. Any possible edge addition will create one or more gems. The lowest order example of this is the butterfly graph, shown in Figure 4.2.

Figure 4.2 Butterfly Graph



Furthermore, this technique does not work on dp graphs. We note that if we add an edge $e = (u, v)$ to a dp graph G with isometric subgraph H where $u \notin V(H)$, or $v \notin V(H)$, or both, the subgraph in $G + e$ induced by $V(H)$ may not be isometric. However, if $u, v \in V(H)$, the subgraph induced by $V(H)$ is isometric in $G + e$, as shown in Lemma 4.2.

Lemma 4.2. *Let G be a graph. Then any set of vertices which induce an isometric subgraph in G will induce an isometric subgraph in $G + e$ if both endpoints of e are in that set.*

Proof. By way of contradiction. Let G be a graph with isometric subgraph $H \subseteq G$, and u and v vertices in $V(H)$ that do not share an edge in G. We will denote $G + uv$ by G' and $H + uv$ by H'. Now assume that w and x are vertices in $V(H')$ which do not have a shortest path in G' consisting entirely of vertices in $V(H')$. Let $P = \{v_1(w), v_2, \ldots, v_{k-1}, v_k(x)\}$ be such a shortest path, with vertices $v_i, \ldots, v_j$ which are not in $V(H')$, where $1 < i < j < k$. Since P cannot contain the edge uv, $d_{G'}(i-1, j+1) = d_G(i-1, j+1)$, and there must be another path P' of the same length as P which contains vertices in $V(H')$ instead of $V_i, \ldots, v_j$. In the event that P' has other vertices not in $V(H')$, we can repeat this process until that is not the case. So w and x do have a shortest path consisting entirely of vertices in $V(H')$, which is a contradiction. Thus H' is an isometric subgraph of G'. □

Now all we need to do is to show that there must exist some edge we can add to an arbitrary dp graph that belongs to an isometric subgraph of almost every order, where isometric subgraphs of the other orders are created by (or at least not destroyed by) the addition of the new edge. We note that adding an edge to a graph will not destroy any isometric subgraphs in which every pair of vertices share an edge or have at least one common neighbor. Lemma 4.3 demonstrates that this is straightforward for any non-complete sequentially dp graph.

Lemma 4.3. *Let G be a sequentially dp graph. If G is not the complete graph, then there exists some new edge e such that $G + e$ is sequentially dp.*

Proof. Let G be a sequentially dp graph on n vertices which is not K_n, with isometric subgraphs $H_1 \subset \ldots \subset H_n$. Let i be the largest integer such that $H_1, \ldots, H_{i-1}$ are cliques in G, and u and v two vertices in H_i which do not share an edge. Now consider $G + uv$ and subgraphs $H_1, \ldots, H_{i-1}, H_i + uv, \ldots, H_n + uv$. $H_1, \ldots, Hi - 1$ are isometric in G' because they are cliques. $H_i, \ldots, H_n$ are isometric in G' by the previous lemma. Since $V(H_i) = V(H_i + uv)$, $V(H_{i+1}) = V(H_{i+1} + uv)$, etc., $H_1 \subset \cdots \subset H_{i-1} \subset H_i + uv \subset \cdots \subset H_n + uv$ as well. Thus $G + e$ is a sequentially dp graph. □

Since all dh graphs are sequentially dp, we can use Lemma 4.3 to add an edge to a dh graph and have the resulting graph be dp. However, we would like to do so and have the resulting graph be dh, which this method does not always do, unlike the one given in Theorem 4.1. We have not been able to apply either of these techniques towards solving the general edge addition problem for dp graphs, stated in Conjecture 4.4.

Conjecture 4.4. *Let G be a non-complete dp graph. Then there exists some new edge e such that $G + e$ is dp.*

We have computationally verified that the following classes of graphs are not counterexamples to the Conjecture 4.4:

- all graphs of order $n \leq 10$

- all regular graphs of order $n \leq 12$

4.2.2 Making Non-DP Graphs DP

Making a non-dh graph into a dh graph is fairly straightforward. We need 2 edges for a domino, 1 edge for a house, 1 edge for a gem, and $n-3$ or $n-4$ edges for an $n-cycle$, depending on whether n is odd or even. This means that for a non-dh graph with entirely edge-disjoint forbidden subgraphs, it is straightforward to count the minimum number of edges needed to make that graph dh. Of course, many non-dh graphs contain overlapping forbidden subgraphs, making the number harder to count. We observe in Table 4.1 that the n-cycle does not require the most edges of all non-dh graphs on n vertices to make it dh except for $5 \leq n \leq 6$.

Table 4.1 Maximum Minimum Number of Additional Edges to Make G a DH Graph

n	size
5	2
6	2
7	5
8	6

When making non-dp graphs into dp graphs it is harder to find a non-naive bound for the number of edges needed. We can always add $n - \Delta(G) - 1$ edges to a maximum degree vertex in a graph G to create a spanning star subgraph. Though adding an edge to a graph may destroy existing isometric subgraphs, this is often not the case, and we see in Table 4.2 that making non-dp graphs into dp graphs takes far fewer edges than with the dh case.

Table 4.2 Maximum Minimum Number of Additional Edges to Make G a DP Graph

n	size
5	1
6	1
7	1
8	1

4.3 Regular Graphs

We start with the case of regular graphs, where the degrees of the vertices are all the same. When we look at Table 4.3, we note that while almost all regular graphs appear to be dp, the convergence is slower than in the case of all graphs. Let (n, r)-regular denote an r-regular graph of order n. Here we only consider *admissible* values of n and r for which (n, r)-regular graphs exist, which is when $n \geq r + 1$ and n and r are not both odd. We want to find a dp (n, r)-regular graph for all admissible values of n and r where a dp graph exists. Note that for $r < 3$, no connected dp (n, r)-regular graph exists, except for very small n. As we saw in Chapter 3, graphs with a high minimum degree are dp, and the union of a number of dp graphs is dp. If connectivity is not a constraint, an algorithm is straightforward.

For the connected case, Ross et al. provide a dp construction for all admissible pairs with $r \geq 3$ [65] in four cases:

- For $n = 2r$
- $n = 2r + 1$
- $n \geq 2r + 2$, $r > 3$
- $n \geq 2r + 2$, $r = 3$

Table 4.3 Percentage of Regular Graphs Which Are DP

n	# connected regular graphs	# connected regular dp graphs	% dp graphs
5	2	1	50.000
6	5	4	80.000
7	4	3	75.000
8	17	14	82.353
9	22	20	90.909
10	167	153	91.617
11	539	484	89.796
12	18979	18405	96.976
13	389436	384319	98.686

Theorem 1. *For each admissible pair (n, r) there exists a dp (n, r) regular graph.*

4.4 Constructing Regular Non-DP Graphs

Some of the more well-known construction methods for generating regular graphs often fail to generate graphs which are dp. Let $C_{n,r}$ be the *circulant* graph, with vertex set $V = \{1, \ldots, n\}$ and edge set E with edge $ij \in E$ if $1 \leq |i-j| \leq r/2$, as well as $i(i+n/2)$ if r is odd, with all calculations here done modulo n. We can prove that (n, r)-regular circulant graphs constructed using an offset list of $\{1, 2, \ldots, \lfloor \frac{r}{2} \rfloor\}$ (with an extra offset of $\frac{n}{2}$ for odd r) are distance preserving when $r \geq \frac{n}{2}$. Since this should be the case for all graphs with $\delta(G) > \frac{n}{2}$ if the Conjecture 3.12 is true, and we have already provided a dp construction for such inputs, we omit the proof here. However, this construction generates non-dp graphs for almost all values of $3 \leq r < \frac{n}{2}$, as seen in Lemmas 4.5 and 4.6.

Lemma 4.5. *For any integers $n \geq 5$ and $2 \leq r < \frac{n}{2}$, where r is even, there exists an r-regular graph on n vertices that is non-dp.*

Proof. Let n and r be integers such that $n \geq 5$ and $2 \leq r < \frac{n}{2}$, where r is even. Then let $G = (V, E)$ be the circulant graph on n vertices with offset list $\{1, \ldots, \frac{r}{2}\}$. Now let $H \subset G$ be an induced subgraph of order $n-1$ whose vertex set is $G \backslash \{v_i\}$, for arbitrary $v_i \in V$. Then $d_G(v_{i-r/2}, v_{i+r/2}) = 2$, and $d_H(v_{i-r/2}, v_{i+r/2}) > 2$, since the only path of length 2 between $v_{i-r/2}$ and $v_{i+r/2}$ in G is the one going through v_i, and the two vertices do not share an edge. So G has no isometric subgraphs of order $n-1$. Thus G is a non-dp graph. □

The circulant graph construction can accommodate even values of n and odd values of r by using the offset $n/2$. We note that the this construction does not yield a non-dp graph when $r = \frac{n}{2} - 1$.

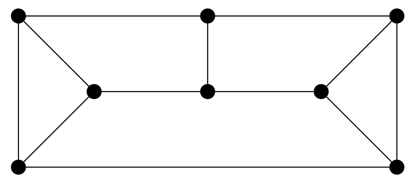
Lemma 4.6. *For any integers $n \geq 5$ and $2 \leq r < \frac{n}{2} - 2$, where n is even and r is odd, there exists an r-regular graph on n vertices that is non-dp.*

Proof. Let n and r be integers such that $n \geq 5$ and $2 \leq r < \frac{n}{2} - 2$, where n is even and r is odd. Then let $G = (V, E)$ be the circulant graph on n vertices with offset list $\{1, \ldots, \frac{r}{2}, \frac{n}{2}\}$. Now let $H \subset G$ be an induced subgraph of order $n-1$ whose vertex set is $G \backslash \{v_i\}$, for arbitrary $v_i \in V$. Then $d_G(v_{i-r/2}, v_{i+r/2}) = 2$, and $d_H(v_{i-r/2}, v_{i+r/2}) > 2$, since the only path of length 2 between $v_{i-r/2}$ and $v_{i+r/2}$ in G is the one going through v_i, and the two vertices do not share an edge. So G has no isometric subgraphs of order $n-1$. Thus G is a non-dp graph. □

Despite the drawbacks of this construction and other similar ones, we have been able to find $(\frac{n}{2} - 1, r)$-regular non-dp graphs for $n = 8$ (see Figure 4.3) and $n = 12$. This leads us to make Conjecture 4.4.

Conjecture 4.7. *For any integers $n \geq 5$ and $2 \leq r < \frac{n}{2}$, where n and r are not both odd, there exists a non-dp (n, r)-regular graph.*

Figure 4.3 A Non-DP $(8,3)$-Regular Graph



4.5 Arbitrary Degree Sequences

We would like to expand our construction abilities to arbitrary sequences of integers which are graphical and have at least one dp construction. In Lemma 4.8 we prove that a modified version of the Havel-Hakimi algorithm generates a dp graph when no reordering of the vertices is done, even when such a reordering would be called for by the original algorithm. However, the modified algorithm fails to generate graphs for many sequences which are graphical.

Lemma 4.8. *Let $S = (d_1, \ldots, d_n)$ be a sequence of n positive integers such that $d_1 \geq \cdots \geq d_n$, where S is a graphical sequence. Let G be the graph constructed using a modified version of the Havel-Hakimi algorithm on S, where no reordering of the sequence / vertices is allowed. If G is constructable using this algorithm, G is a dp graph, with the set of isometric subgraphs $\{H_1, \ldots, H_n\}$, where $H_k \in \{H_1, \ldots, H_n\}$ is the graph induced by the vertices corresponding to the first k terms of S.*

Proof. Proof is by induction on n.

Basis. For $n = 1$, the only graphical sequence is (0), which is constructable using the modified Havel-Hakimi algorithm, and $G = K_1$ with the isometric subgraph $H_1 = G$ is dp.

Inductive Hypothesis. Assume the statement holds true for $n = 1, 2, \ldots, N$.

Inductive Step. For the graphical sequence $S = (d_1, \ldots, d_{N+1})$, we construct G from S using the modified Havel-Hakimi algorithm. Let $V(G) = \{v_1, \ldots, v_{N+1}\}$, where vertex $v_i \in V(G)$ corresponds to $d_i \in S$. Let $G' = G \backslash \{v_{N+1}\}$. By the inductive hypothesis, G' is dp, with isometric subgraphs $\{H_1, \ldots, H_N\}$. Adding back the vertex v_{N+1}, we consider an arbitrary $H_k \in \{H_1, \ldots, H_N\}$ as a subgraph of G. Let w and x be arbitrary vertices such that $w \leq x \leq k$. We want any shortest path P between w and x in H_k to also be shortest in G. This must be the case, since a shortest path involving v_{N+1} would necessitate w or one of the intermediate vertices in P sharing an edge directly with x, and P would still be a shortest path. So $\{H_1, \ldots, H_N, G\}$ is a complete set of isometric subgraphs of G.

Thus G is a dp graph. □

Table 4.4 Success Rate of the Modified Havel-Hakimi Algorithm

n	# graphical degree sequences	# successes	% successes
5	20	12	60.000
6	71	32	45.070
7	240	86	35.833
8	871	243	27.899
9	3149	703	22.332
10	11655	2094	17.967
11	43332	6369	14.698
12	162769	19770	12.146

Even if we continue without reordering the vertices when such a reordering is needed, the algorithm will still often fail. In Table 4.4 the results of the modified Havel-Hakimi algorithm are provided for $5 \leq n \leq 12$. Note that any time the algorithm terminates properly is counted as a success, whether the resulting graph is connected or not, as some graphical degree sequences do not have a connected representation.

4.6 Summary

Constructing connected dp graphs is difficult for many constraints, in particular when dealing with an arbitrary degree sequence. Although we have an algorithm that works for certain cases, we wish to find one that works for all degree sequences. Also of interest is the problem of determining the minimum number of edges needed to make a non-distance preserving graph into a dp one. While we do not have any conjecture here that improves on the trivial $|V(G)| - \delta(G) - 1$ upper bound, experimental evidence indicates that this number is quite low. The question here is how much of an improvement we can make over augmenting the graph so that it contains a star subgraph with the least number of additional edges.

Chapter 5

Isometric Subgraphs

More efficient isometric subgraph finding is integral to recognizing dp graphs. Our computational work in Chapters 3 and 4 used brute forced methods, which is slow even for very low order graphs. In practice we are often able to find quite large isometric subgraphs using Monte Carlo methods. This is despite our observations that the larger a graph is, the lower the ratio of isometric to non-isometric subgraphs tends to be.

5.1 Finding Isometric Subgraphs

Lemma 3.18, along with a proof of Conjecture 3.19, would demonstrate that the recognition problem for dp graphs is NP-Complete. Our experiences so far suggest that it is, and we have commenced searching for sub-isometric subgraph finding heuristics under this assumption. The first observation here is that an isometric subgraph of order k is not necessarily constructable from some isometric subgraph of order $k-1$. However, we can still construct an incremental Monte Carlo algorithm that attempts to find an isometric subgraph in G of up to order k:

i. Select an arbitrary vertex from G. This vertex is trivially an isometric subgraph of order 1, which we denote H.

ii. Attempt to add a neighbor from G not in H to H such that the resulting induced subgraph is dp. If no such vertex exists, stop.

iii. If H is of order k, stop. Otherwise, go back to step ii.

While inelegant, randomized, and computationally inefficient, this algorithm may be used to find reasonably high order isometric subgraphs. When we need to be certain whether a graph contains an isometric subgraph of order k, we use a brute force search. For experimental results, see our previous paper [59].

5.2 Bounds on the Number of Isometric Subgraphs

Tables 5.1 and 5.2 illustrate the potential difficulty of finding isometric subgraphs. While the average number of isometric subgraphs increases with the order of a graph, the percentage of all induced subgraphs which are isometric decreases.

Table 5.1 Average Number of Isometric Subgraphs of All Connected Graphs

$n\backslash k$	1	2	3	4	5	6	7	8	9
1	1.000								
2	2.000	1.000							
3	3.000	2.500	1.000						
4	4.000	4.167	3.333	1.000					
5	5.000	6.190	6.952	3.857	1.000				
6	6.000	8.491	12.179	9.839	4.958	1.000			
7	7.000	11.198	19.355	19.60	12.907	5.190	1.000		
8	8.000	14.412	29.284	35.167	29.011	16.322	5.780	1.000	
9	9.000	18.219	42.782	58.951	57.802	40.825	20.173	6.382	1.000

Table 5.2 Average Percentage of Isometric Subgraphs to Total Subgraphs of all Connected Graphs

$n\backslash k$	1	2	3	4	5	6	7	8	9
1	100.0								
2	100.0	100.0							
3	100.0	83.3	100.0						
4	100.0	69.4	83.3	100.0					
5	100.0	61.9	69.5	77.1	100.0				
6	100.0	56.6	60.9	65.6	76.6	100.0			
7	100.0	53.3	55.3	56.0	61.5	74.1	100.0		
8	100.0	51.5	52.3	50.2	51.8	58.3	72.2	100.0	
9	100.0	50.6	50.9	46.8	45.9	48.6	56.0	70.9	100.0

5.3 Applications

In [59], we proposed a clustering algorithm using isometric subgraphs [59]. While the results of our algorithm compare favorably to hierarchical clustering methods, we need a more sophisticated algorithm for finding isometric subgraphs. This is because of the poor time complexity of the current one. In addition to using the incremental isometric subgraph algorithm described in Section 5.1, it introduces the notion of almost dp subgraphs. For a subgraph $H \subseteq G$, we define the *average distance increase* for H as the sum of the distance increases between H and G divided by the number of vertices in H. If H is actually an isometric subgraph of G it will have an average distance increase of 0. This relaxation allows us to address the fact that we are not always able to disjointly partition G into an arbitrary number of isometric subgraphs.

5.4 Summary

Our work in [59] and on small Twitter datasets demonstrates that isometric or nearly isometric subgraphs do make good clusters, at least for certain kinds of datasets. However, a practical algorithm will require not using exact distances, as the all pairs shortest paths problem requires $O(VE+v^2\ log\ V)$ (Bellman-Ford) to $O(V^3)$ (Floyd-Warshall) time. Even so, a better than brute force non-approximation algorithm would still be helpful for theoretical exploration and practical problems involving small datasets.

Chapter 6

Related Work

Hundreds of graph classes have been covered in the literature. Many families of graphs, including such well known ones as trees and bipartite graphs, form a large hierarchical structure, with perfect graphs at the top. In this chapter we will review the literature on perfect graphs and relevant subclasses, particularly dh graphs.

Before we proceed, we need to define some standard functions on graphs. The following definitions and notation are taken from Brandstädt et al. [8]. Let $G = (V, E)$ be a graph.

- The *clique number* of G, denoted $\omega(G)$, is the order of the largest clique in G. A *clique* is a subset of vertices such that every pair of vertices in the subset are adjacent.

- The *chromatic number* of G, denoted $\chi(G)$, is the minimum number of independent sets V can be partitioned into. An *independent set* is a subset of vertices such that no two vertices in the subset are adjacent.

- The *stability number* of G, denoted $\alpha(G)$, is the order of the largest independent (stable) set in G.

- The *clique cover number* of G, denoted $k(G)$, is the minimum number of cliques V can be partitioned into. Note that we use $k(G)$ rather than $\kappa(G)$ here, as $\kappa(G)$ is used to denote vertex connectivity in Bondy and Murty [7], and elsewhere.

All four of these problems are included in Karp's list of 21 NP-Complete problems [48]. Since the complement of a clique is an independent set, the equalities $\alpha(G) = \omega(\overline{G})$ and $k(G) = \chi(\overline{G})$ immediately follow from the above definitions [35]. Furthermore, a clique and an independent set can share at most one vertex, so we have the inequalities $\omega(G) \leq \chi(G)$ and $\alpha(G) \leq k(G)$ as well [35].

6.1 Perfect Graphs

First formalized by Berge in the 1960's [4, 6], a *perfect graph* is one where for every subset $A \subseteq V$, we have $\chi(G[A]) = \omega(G[A])$, where $G[A]$ is the subgraph of G induced by A. In the context of perfect graphs and related graph classes, an induced cycle of order 5 or more is often referred to as a *hole*. An *antihole*, denoted $\overline{C_n}$, is the complement of a hole of order n. This definition can be reformulated in a number of ways using the previously stated identities. Berge also provided us with two important conjectures (now theorems) regarding perfect graphs [4, 5]:

- **(Weak) Perfect Graph Conjecture.** The complement of a perfect graph is perfect.
- **Strong Perfect Graph Conjecture.** A graph is perfect if and only if it does not contain any odd order holes or odd order antiholes.

Obviously, the latter statement implies the former. These theorems and other nontrivial characterizations are discussed below.

6.1.1 Characterizations

The following characterizations of perfect graphs are equivalent:

i. $\omega(G[A]) = \chi(G[A])$, for all $A \subseteq V$.

ii. $\alpha(G[A]) = k(G[A])$, for all $A \subseteq V$.

iii. $\omega(G[A]) \cdot \alpha(G[A]) \geq |A|$, for all $A \subseteq V$.

iv. $P_I(A) = P(A)$, where A is the clique matrix of G. The *clique matrix* of G is the maximal cliques versus vertices matrix.

v. G contains no odd holes or antiholes.

Characterization (i) is Berge's definition of a perfect graph. In 1972, Lovász demonstrated the equivalence of (i) and (ii), proving the Perfect Graph Conjecture, now known as the Perfect Graph Theorem [52]. In another paper Lovász sharpened this result by proving the further equivalence of (iii), although this characterization is still weaker than the Strong Perfect Graph Conjecture [51]. Chvátal gave the polyhedral characterization of (iv) in 1975 [18]. The equivalence of characterization (v) and (i) is the Strong Perfect Graph Conjecture. In 2005 it was finally proven by Chudnovsky et al., and is now known as the Strong Perfect Graph Theorem [16]. Subclasses of perfect graphs include bipartite graphs, dh graphs and trees [8].

6.1.2 Strong Perfect Graph Theorem

Minimal imperfect graphs appear often in attempts to prove the Strong Perfect Graph Theorem. A graph is *minimal imperfect* if it is not a perfect graph, but every proper induced subgraph is a perfect graph [35]. The Strong Perfect Graph Theorem can be restated in terms of minimal imperfect graphs: The only minimal imperfect graphs are odd holes and odd antiholes [8]. The following are properties of minimal imperfect graphs:

i. $|V| = \alpha(G)\omega(G) + 1$ [35].

ii. Every vertex in G belongs to exactly $\omega(G)$ maximum cliques of size $\omega(G)$ [60].

iii. Every vertex in G belongs to exactly $\alpha(G)$ maximum independent sets of size $\alpha(G)$ [60].

iv. G contains exactly $|V|$ maximum cliques of size $\omega(G)$ [60].

v. G contains exactly $|V|$ maximum independent sets of size $\alpha(G)$ [60].

vi. G does not contain a star-cutset [19]. A *star-cutset* is a set of vertices $X \subset V(G)$ whose removal disconnects G, and that there exists some vertex $u \in V$ adjacent to every other vertex in X.

vii. G does not contain an even pair [55].

viii. G is $(2\omega(G) - 2)$-connected [68].

For a deeper analysis of attempts to prove the Strong Perfect Graph Theorem, see Rousse et al. [66].

6.1.3 Complexity

For some time all that was known about perfect graph recognition was that the problem of recognizing Berge graphs belonged to Co-NP [53]. It has recently been shown that Berge graphs, and hence perfect graphs, may be recognized in polynomial time [15]. For perfect graphs, ordinarily NP-Complete problems such as the chromatic number, clique problem, and independent set problem are solvable in polynomial time [54].

6.2 Distance-Hereditary Graphs

First proposed by Howorka [43], a *dh* graph is one in which every connected induced subgraph is isometric. Given a a connected graph G with interval function

$$I(u, v) = \{x \mid x \text{ is a vertex of } G \text{ on some shortest } (u, v)\text{-path}\},$$

Bandelt and Mulder provide a number of equivalent characterizations:

i. G is distance hereditary.

ii. For any two vertices u and v with $d_G(u,v) = 2$, there is no induced $(u,v)-$path of length greater than 2.

iii. The house, hole (long induced cycle), domino, and gem are not induced subgraphs of G.

iv. The house, hole (long induced cycle), domino, and gem are not isometric subgraphs of G.

v. The house, domino, and gem are not induced (or isometric) subgraphs of G, and $I(u,v) \cap I(u,w) = \{v\}$ implies $d_G(u,w) \geq d_G(u,v) + d_G(v,w) - 1$.

vi. The gem is not an induced subgraph of G, and for any three vertices u, v, w, at least two of the following inclusions hold: $I(u,v) \subseteq I(u,w) \cup I(v,w)$, $I(u,w) \subseteq I(u,v) \cup I(v,w)$, and $I(v,w) \subseteq I(u,v) \cup I(u,w)$.

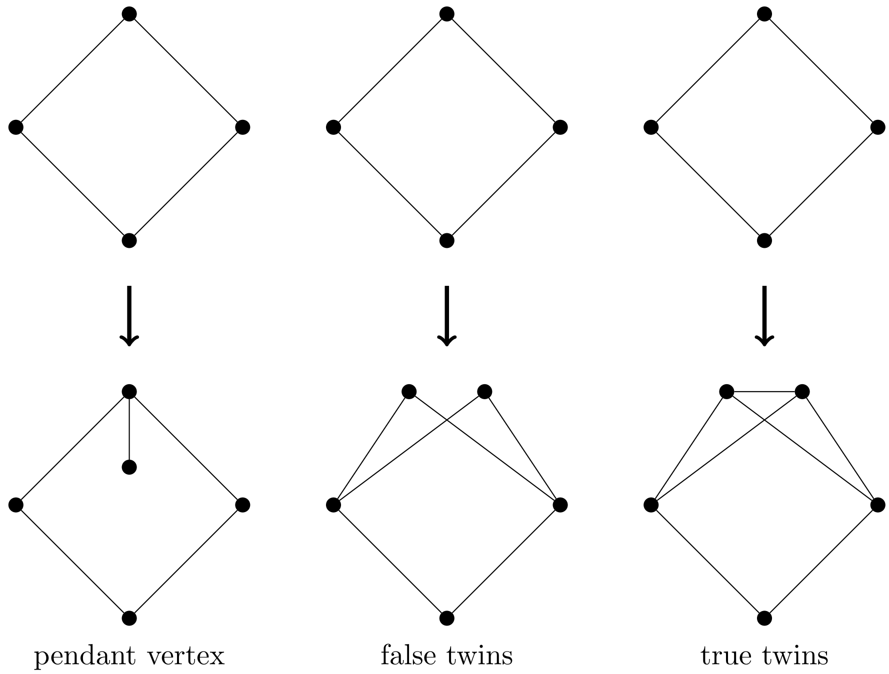
vii. For any four vertices u, v, w, x, at least two of the following distance sums are equal: $d_G(u,v) + d_G(w,x)$, $d_G(u,w) + d_G(v,x)$, $d_G(u,x) + d_G(v,w)$.

viii. G satisfies condition (vii), and if in (vii) the smaller distance sums are equal, then the largest one exceeds the smaller ones by at most 2.

Dh graphs are a subset of perfect graphs [43].

Hammer and Maffrey proposed a linear time recognition algorithm [39]. This algorithm was later shown to be incorrect by Damiand et al., who provide their own linear time recognition algorithm that attemps to decompose a graph into a sequence of pendant vertex and twin operations [22], shown in Figure 6.1. As a subset of several other graph classes, dh graphs inherit a number of polynomial time optimization algorithms. Several algorithms designed specifically for dh graphs also exist, including a linear time algorithm for the Hamiltonian cycle problem [45].

Figure 6.1 Pendant Vertex and Twin Operations



6.3 Geodetically Connected Graphs

The connectivity of a non-complete graph G is the number of vertices required to disconnect the graph. Similarly, the geodetic connectivity (gc) of G is the number of vertices required to increase $d_G(u, v)$ for some $u, v \in V(G)$. First defined by Entringer et al. [28], these graphs are also known as self-repairing graphs [32, 33]. While not as well studied as perfect or dh graphs, a number of other papers attempt to characterize [27, 56] and find minimum examples [57, 62, 63, 50] of gc graphs.

Chang and Ho provide a polynomial time recognition algorithm for geodetically connected graphs [11], and a linear time recognition algorithm for some specific cases [14, 12]. Other papers of Chang examine special cases of 2-gc graphs, which are referred to as hinge-free graphs [13, 10]. Let G be a graph and $k \geq 2$ be an integer. Then the following characterizations are equivalent [13]

i. G is a $k-GC$ graph.

ii. Every pair of nonadjacent vertices in G are joined by at least k vertex-disjoint geodesics.

iii. Every pair of nonadjacent vertices in G are joined by at least k edge-disjoint geodesics.

iv. G is a $k-GEC$ graph.

v. Every pair of vertices $u, v \in V(G)$ with $d(u, v) = 2$ satisfies $|N(u) \cap N(v)| \geq k$.

6.4 Distance-Preserving Graphs

There are a number of concepts in the graph theory literature that involve constraints on distances or other invariants as vertices are deleted from the graph. Except for the work on gc graphs, they are subtly but significantly different that dp graphs and isometric subgraphs, and their proof techniques have not been applicable to our work. Recently, Zahedi proved several characterizations and constructions of dp graphs [70]. Notably, chordal graphs are dp. Also, every graph with girth of at least 5 where each vertex is a cut vertex or contained in a cycle is non-dp.

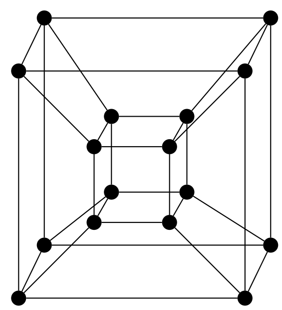
6.5 Other Graph Classes

For further coverage of perfect graphs and related subclasses, including dh graphs, see Brandstädt et al. [8], Alfonsin and Reed [1], and Golumbic [35]. In this section we cover several other graph classes with connections to dp graphs.

6.5.1 Hypercubes

The n-dimensional hypercube, denoted Q_n, is the graph on 2^n vertices whose vertex set is the set of all n-tuples of 0s and 1s, where two vertices share an edge if their n-tuples differ by exactly one element [7]. While they are usually referred to as hypercubes, or n-cubes, they previously went by the term measure polytopes [21]. Less formally, an n-cube is a generalization of the cube to n dimensions. They can also be defined recursively using the Cartesian product, where $Q_1 = K_2$, and $Q_n = K_2 \times Q_{n-1}$ [40]. Figure 6.2 gives a drawing of Q_4.

Figure 6.2 A Tesseract (Q_4)



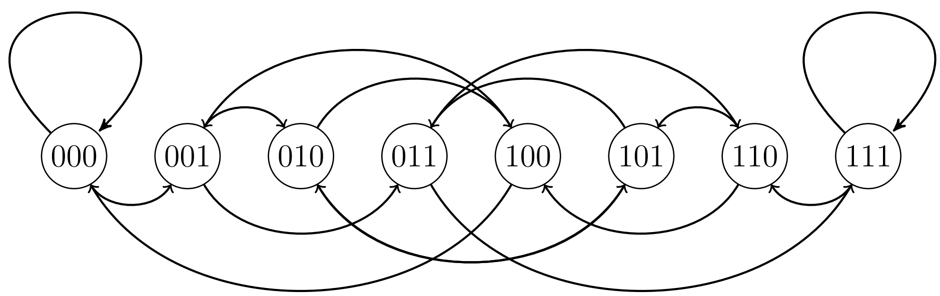
Given the nature of the tuples associated with each vertex, the distance between any two vertices is the number of elements that differ between their respective tuples. The average distance between two vertices in Q_n is $\frac{n \cdot 2^{n-1}}{2^n - 1}$, which approaches $n/2$ as n goes to infinity [47]. The vertex connectivity and edge connectivity of Q_n is n, i.e., Q_n cannot be disconnected without removing a minimum of n vertices or edges [41].

6.5.2 De Bruijn Digraphs

An n-dimensional de Bruijn digraph, denoted $B_{m,n}$, is the directed graph on m^n vertices, where the vertex set is the set of all sequences of length n over an alphabet of cardinality m, and there exists an arc from vertex $(a_1 a_2 \cdots a_n)$ to vertex $(b_1 b_2 \cdots b_n)$ if and only if $a_{i+1} = b_i$ for $1 \le i \le n-1$ [7]. In other words, vertex $(a_1 a_2 \cdots a_n)$ has arcs going to vertices $(a_2 a_3 \cdots a_n *)$, where $*$ is an arbitrary element in the alphabet. Also known as de Bruijn-Good digraphs, they were discovered independently by de Bruijn [23] and Good [36]. Like hypercubes, de Bruijn digraphs have applications in communication and multiprocessor networks [67]. They have also been used to represent genomic sequences in bioinformatics [61, 71]

De Bruijn digraphs may be generalized as simple directed graphs by removing multiple edges and loops [46]. The undirected version of the generalized de Bruijn graph further replaces arcs with edges [64, 30]. We consider undirected generalized de Bruijn graphs over the alphabet $\{0, 1\}$, as shown in Figure 6.3.

Figure 6.3 A De Bruijn Digraph ($B_{2,3}$)



Chapter 7

Conclusions

In this work we defined dp graphs, a fundamentally new class class of graphs. We proved several characterizations involving dp graphs, and provided construction methods for a variety of constraints. We proposed a number of key conjectures critical to the understanding of dp graphs. Additionally, we used a simple isometric subgraph finding algorithm to demonstrate that dp graphs and isometric subgraphs have some applications in data mining. It is our hope that future researchers will not only solve these conjectures, but find more applications for dp graphs as well, by using what we have learned.

APPENDICES

APPENDIX A LIST OF SYMBOLS

$V(G)$	vertex set of G
$E(G)$	edge set of G
ψ_G	incidence function of G
$\delta(G)$	minimum vertex degree of G
$\Delta(G)$	maximum vertex degree of G
$G[X]$	subgraph of G induced by X
$\overline{G}$	complement of G
$d_G(u, v)$	distance function of G
P_n	path of order n
C_n	cycle of order n
S_n	star graph of order n
K_n	complete graph of order n
$K_{m,n}$	complete digraph with partitions of order m and n
$\omega(G)$	clique number of G
$\chi(G)$	chromatic number of G
$\alpha(G)$	stability number of G

$k(G)$	clique cover number of G
Q_n	n-dimensional hypercube of order 2^n
$B_{m,n}$	n-dimensional De Bruijn graph of order m^n

APPENDIX B SPECIAL GRAPHS

Figure B.1 Butterfly Graph

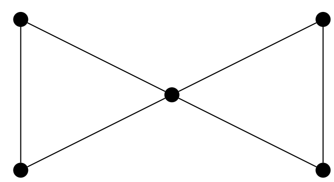


Figure B.2 Complete (K_n)

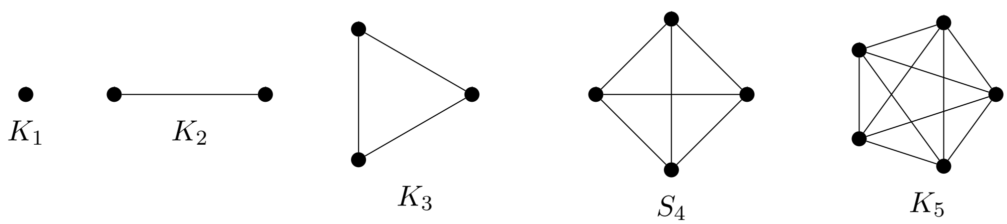


Figure B.3 Complete Bipartite ($K_{m,n}$)

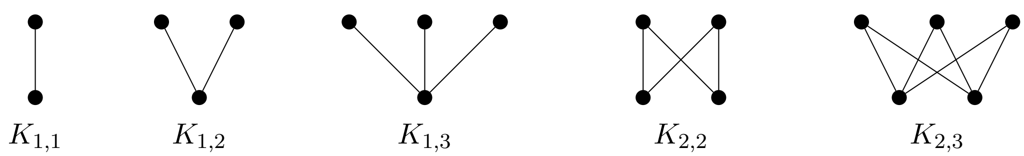


Figure B.4 Cycle (C_n)

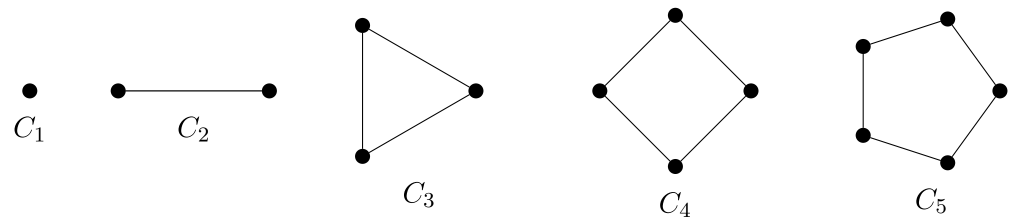


Figure B.5 Domino

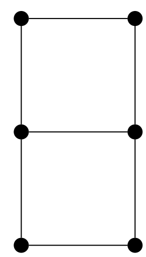


Figure B.6 Gem

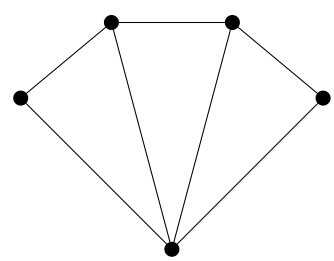


Figure B.7 House

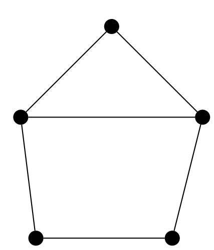


Figure B.8 Hypercube (Q_n)

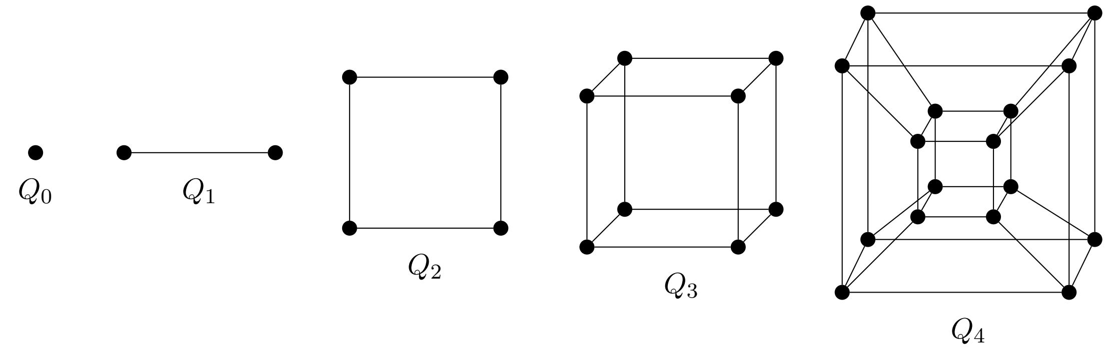


Figure B.9 Pan

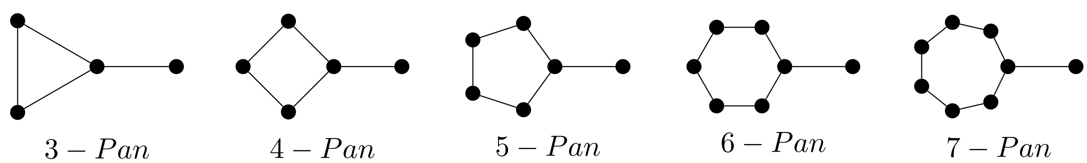


Figure B.10 Path (P_n)

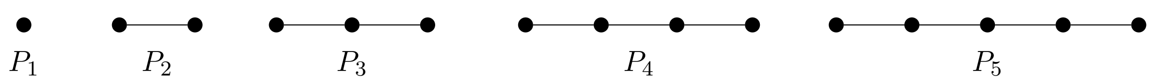
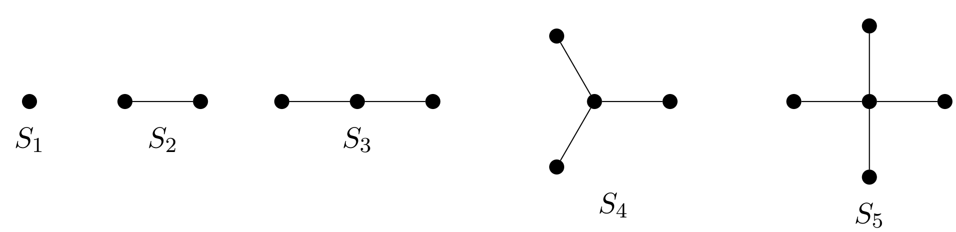


Figure B.11 Star (S_n)



APPENDIX C REAL WORLD NETWORKS

Politics (UK), Politics (Ireland), Premier League Football, Olympics 2012, and Rugby Union are political and sports related Twitter datasets. A node in these datasets represents a Twitter user. A link between two nodes represents one or both of the users following the other. The class labels are political parties, sports teams, or in the case of Olympics 2012, sports.

The CiteSeer and Cora datasets are paper citation networks. A node in these datasets represents a scientific publication. A link between two nodes represents a citation by one of the papers of the other. The class labels are areas of research. The original CiteSeer dataset has 3312 nodes and 4536 links. We extracted the largest component, which has 2110 nodes and 3668 links. The original Cora dataset has 2708 nodes and 5378 links. We extracted the largest component, which has 2485 nodes and 5069 links.

Tables C.1 through C.7 consider distances in the clusters formed by the class labels for these seven datasets. The distances between vertices in the subgraphs induced by the class labels are compared to the distances between those vertices in the original graph. Tables C.8 and C.9 examine the same metrics using clusters found with k-means and hierarchical clustering algorithms instead of those based on the class labels.

Table C.1 UK MPs on Twitter (419 nodes, 27340 links, 5 communities)

class	order	avg. distance	avg. distance increase	# infinite paths
conservative	173	1.58	0.0	0
labour	187	1.44	0.0	186
libdem	43	1.25	0.0	0
snp	5	1.0	0.0	0
other	11	1.62	0.28	26

Table C.2 Irish Politicians and Organizations on Twitter (348 nodes, 16856 links, 7 communities)

class	order	avg. distance	avg. distance increase	# infinite paths
ff	49	1.39	0.0	0
fg	143	1.56	0.0	0
green	7	1.0	0.0	0
ind	31	1.71	0.04	0
labour	79	1.28	0.0	0
sf	31	1.31	0.0	0
ula	8	1.14	0.0	0

Table C.3 Premier League Players and Clubs on Twitter (248 nodes, 3819 links, 20 communities)

class	order	avg. distance	avg. distance increase	# infinite paths
arsenal	12	1.36	0.0	0
aston-villa	11	1.2	0.0	0
chelsea	12	1.23	0.0	0
everton	15	1.36	0.0	0
fulham	10	1.16	0.0	0
liverpool	13	1.24	0.0	0
man-city	10	1.13	0.0	0
man-utd	10	1.18	0.0	0
newcastle	10	1.31	0.0	0
norwich	9	1.36	0.0	0
qpr	10	1.4	0.0	0
reading	14	1.07	0.0	0
southampton	14	1.24	0.0	0
spurs	23	1.47	0.0	0
stoke	15	1.43	0.0	0
sunderland	13	1.13	0.0	0
swansea	11	1.07	0.0	0
west-brom	17	1.26	0.0	31
west-ham	8	1.29	0.04	0
wigan	11	1.25	0.0	0

Table C.4 Olympic Athletes and Organizations on Twitter (464 nodes, 10642 links, 28 communities)

class	order	avg. distance	avg. distance increase	# infinite paths
archery	7	1.14	0.0	0
athletics	52	1.53	0.0	0
badminton	13	1.26	0.0	0
basketball	14	1.27	0.0	0
beach-volleyball	9	1.39	0.0	0
boxing	21	1.19	0.0	0
canoeing	19	1.23	0.0	0
cycling	29	1.39	0.0	0
diving	20	1.31	0.0	0
equestrianism	15	1.54	0.0	0
fencing	21	1.45	0.0	0
gymnastics	16	1.38	0.0	0
handball	13	1.1	0.0	0
hockey	43	1.3	0.0	0
judo	14	1.32	0.0	0
pentathlon	11	1.02	0.0	0
rowing	21	1.3	0.0	0
sailing	10	1.33	0.0	0
shooting	7	1.14	0.0	0
swimming	34	1.19	0.0	0
swimming-sync	10	1.0	0.0	0
tabletennis	7	1.24	0.0	0
taekwondo	11	1.18	0.0	10
tennis	12	1.67	0.0	11
triathlon	7	1.0	0.0	0
waterpolo	20	1.12	0.0	0

Table C.4 (cont'd)

class	order	avg. distance	avg. distance increase	# infinite paths
weightlifting	3	1.33	0.0	0
wrestling	5	1.2	0.0	0

Table C.5 Rugby Players and Clubs on Twitter (854 nodes, 35757 links, 15 communities)

class	order	avg. distance	avg. distance increase	# infinite paths
america	2	1.0	0.0	0
argentina	1	—	—	0
australia	61	1.53	0.0	0
canada	1	—	—	0
england	218	1.86	0.01	433
fiji	2	1.0	0.0	0
france	105	1.8	0.02	0
ireland	91	1.6	0.01	90
italy	18	1.48	0.0	0
new-zealand	64	1.6	0.0	186
samoa	11	1.25	0.0	0
scotland	63	1.44	0.0	0
south-africa	99	1.77	0.01	480
tonga	5	1.33	0.0	7
wales	113	1.62	0.01	112

Table C.6 CiteSeer (2110 nodes, 3668 links, 6 communities)

class	order	avg. distance	avg. distance increase	# infinite paths
Agents	463	5.56	0.42	33007
IR	532	4.57	0.28	44185
DB	388	8.59	0.6	42500
AI	115	1.64	0.08	6455
HCI	304	7.42	0.31	25056
ML	308	5.89	0.07	37188

Table C.7 Cora (2485 nodes, 5069 links, 7 communities)

class	order	avg. distance	avg. distance increase	# inf. paths
Neural Networks	726	5.72	0.49	61201
Rule Learning	131	3.77	0.27	3338
Reinforcement Learning	214	3.26	0.08	7190
Probabilistic Methods	379	6.41	0.32	16282
Theory	344	4.72	0.59	17892
Genetic Algorithms	406	3.54	0.01	4396
Case Based	285	5.27	0.72	11034

Table C.8 *k*-means Clustering

dataset	# trials	cluster order	cluster order σ	average distance	average dist. increase	average # inf. paths	average entropy
Politics (UK)	100	83.60	28.63	1.38	0.19	305.49	0.16
Politics (IE)	100	49.71	19.47	1.41	0.08	156.18	0.34
Football	100	12.41	6.21	1.51	0.15	14.95	0.72
Olympics	100	16.67	9.36	1.50	0.20	34.90	0.61
Rugby	100	56.53	26.52	1.70	0.17	85.56	0.48
CiteSeer	20	351.67	173.01	4.89	0.02	21328.48	0.69
Cora	20	355.00	136.61	4.16	0.70	29814.84	0.67

Table C.9 Hierarchical Clustering (Average Linkage)

dataset	cluster order	cluster order σ	average distance	average dist. increase	average # infinite paths	average entropy
Politics (UK)	83.60	161.20	1.78	0.00	1.80	0.65
Politics (IE)	58.00	105.99	1.77	0.00	0.50	0.72
Football	12.35	6.78	1.47	0.01	0.00	0.53
Olympics	16.54	16.91	1.55	0.02	0.00	0.43
Rugby	56.63	62.60	1.77	0.01	0.00	0.39
CiteSeer	351.67	718.84	8.84	0.00	0.00	0.91
Cora	355.00	803.11	5.90	0.00	0.00	0.89

REFERENCES

REFERENCES

[1] J.L.R. Alfonsin and B.A. Reed. *Perfect graphs.* John Wiley & Sons Inc, 2001.

[2] L.W. Beineke. Derived graphs and digraphs. *Beiträge zur Graphentheorie*, pages 17–33, 1968.

[3] L.W. Beineke. Characterizations of derived graphs. *Journal of Combinatorial Theory*, 9(2):129–135, 1970.

[4] C. Berge. Färbung von Graphen, deren sämtliche bzw. deren ungerade Kreise starr sind. *Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe*, 10:114, 1961.

[5] C. Berge. Sur une conjecture relative au probleme des codes optimaux, commun. 13 eme assemblee gen. *URSI, Tokyo*, 1962.

[6] C. Berge. Some classes of perfect graphs. *Six Papers on Graph Theory*, pages 1–21, 1963.

[7] J.A. Bondy and U.S.R. Murty. *Graph Theory (Graduate Texts in Mathematics).* Springer Berlin, 2008.

[8] A. Brandstädt and J.P. Spinrad. *Graph classes: a survey.* Society for Industrial Mathematics, 1999.

[9] G. Brinkmann, K. Coolsaet, J. Goedgebeur, and H. Mélot. House of graphs: A database of interesting graphs. *Discrete Applied Mathematics*, 161(1):311–314, 2013.

[10] J-M. Chang. *Algorithmic Aspects of Some Geodetic Problems in Special Graphs.* PhD thesis, National Central University, 2001.

[11] J-M. Chang and C-W. Ho. The recognition of geodetically connected graphs. *Information Processing Letters*, 65(2):81–88, 1998.

[12] J-M. Chang and C-W. Ho. Recognizing hinge-free line graphs and total graphs. *Taiwanese Journal of Mathematics*, 5(4):pp–789, 2001.

[13] J-M. Chang, C-W. Ho, C-C. Hsu, and Y-L. Wang. The characterizations of hinge-free networks. *Proceeding of International Computer Symposium on Algorithms*, pages 105–112, 1996.

[14] J-M. Chang, C-C. Hsu, Y-L. Wang, and T-Y. Ho. Finding the set of all hinge vertices for strongly chordal graphs in linear time. *Information sciences*, 99(3):173–182, 1997.

[15] M. Chudnovsky, G. Cornuéjols, X. Liu, P. Seymour, and K. Vušković. Recognizing berge graphs. *Combinatorica*, 25(2):143–186, 2005.

[16] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas. The strong perfect graph theorem. *Annals of Mathematics*, pages 51–229, 2006.

[17] V. Chvátal. Set-packing problems and threshold graphs, 1973.

[18] V. Chvátal. On certain polytopes associated with graphs. *Journal of Combinatorial Theory, Series B*, 18(2):138–154, 1975.

[19] V. Chvátal. Star-cutsets and perfect graphs. *Journal of Combinatorial Theory, Series B*, 39(3):189–199, 1985.

[20] D.G. Corneil, Y. Perl, and L.K. Stewart. A linear recognition algorithm for cographs. *SIAM Journal on Computing*, 14(4):926–934, 1985.

[21] H.S.M. Coxeter and H.S.M. Coxeter. *Regular polytopes*. Dover Publications, 1973.

[22] G. Damiand, M. Habib, and C. Paul. A simple paradigm for graph recognition: application to cographs and distance hereditary graphs. *Theoretical Computer Science*, 263(1-2):99–111, 2001.

[23] N.G. de Bruijn and P. Erdős. A combinatorial problem. *Koninklijke Nederlandse Akademie v. Wetenschappen*, 49(49):758–764, 1946.

[24] D.G. Degiorgi and K. Simon. A dynamic algorithm for line graph recognition. In *Graph-Theoretic Concepts in Computer Science*, pages 37–48. Springer, 1995.

[25] R. Diestel. *Graph Theory (Graduate Texts in Mathematics)*. Springer-Verlag Berlin and Heidelberg GmbH & Company KG, 2005.

[26] G.A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952.

[27] H. Enomoto and A. Saito. Disjoint shortest paths in graphs. *Combinatorica*, 4(4):275–279, 1984.

[28] R. Entringer, D. Jackson, and P. Slater. Geodetic connectivity of graphs. *Circuits and Systems, IEEE Transactions on*, 24(8):460–463, 1977.

[29] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5:17–61, 1960.

[30] A-H. Esfahanian and S.L. Hakimi. Fault-Tolerant Routing in DeBruijn Comrnunication Networks. *IEEE Transactions on Computers*, pages 777–788, 1985.

[31] A-H. Esfahanian and O.R. Oellermann. Distance-hereditary graphs and multidestination message-routing in multicomputers. *Journal of Comb. Math. and Comb. Computing*, 13:213–222, 1993.

[32] A.M. Farley and A. Proskurowski. Self-repairing networks. *Parallel Processing Letters*, 3(04):381–391, 1993.

[33] A.M. Farley and A. Proskurowski. Minimum self-repairing graphs. *Graphs and Combinatorics-an Asian Journal*, 13(4):345–352, 1997.

[34] M.C. Golumbic. Trivially perfect graphs. *Discrete Mathematics*, 24(1):105–107, 1978.

[35] M.C. Golumbic. *Algorithmic graph theory and perfect graphs*. North-Holland, 2004.

[36] I.J. Good. Normal recurring decimals. *Journal of the London Mathematical Society*, 1(3):167, 1946.

[37] D. Greene and P. Cunningham. Producing a unified graph representation from multiple social network views. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 118–121. ACM, 2013.

[38] G. Grimmett. *What is Percolation?* Springer, 1999.

[39] P.L. Hammer and F. Maffray. Completely Separable Graphs*. *Discrete Applied Mathematics*, 27(1-2):85–99, 1990.

[40] F. Harary. Graph Theory. 1969.

[41] F. Harary, J.P. Hayes, and H.J. Wu. A survey of the theory of hypercube graphs. *Computers & Mathematics with Applications*, 15(4):277–289, 1988.

[42] F. Harary and R.Z. Norman. Some properties of line digraphs. *Rendiconti del Circolo Matematico di Palermo*, 9(2):161–168, 1960.

[43] E. Howorka. A characterization of distance-hereditary graphs. *The Quarterly Journal of Mathematics*, 28(4):417, 1977.

[44] E. Howorka. A characterization of ptolemaic graphs. *Journal of Graph Theory*, 5(3):323–331, 1981.

[45] R.W. Hung and M.S. Chang. Solving the path cover problem on circular-arc graphs by using an approximation algorithm. *Discrete Applied Mathematics*, 154(1):76–105, 2006.

[46] M. Imase and M. Itoh. A design for directed graphs with minimum diameter. *Computers, IEEE Transactions on*, 100(8):782–784, 1983.

[47] P.C. Kainen. A lower bound for crossing numbers of graphs with applications to Kn, Kp, q, and Q (d). *Journal of Combinatorial Theory, Series B*, 12(3):287–298, 1972.

[48] R.M. Karp. Reducibility Among Combinatorial Problems. *Complexity of computer computations*, pages 85–103, 1972.

[49] D.C. Kay and G. Chartrand. A characterization of certain ptolemaic graphs. *Canad. J. Math*, 17:342–346, 1965.

[50] Y. Lan and S. Chen. Some special minimum k-geodetically connected graphs. *Discrete Applied Mathematics*, 159(10):1002–1012, 2011.

[51] L. Lovász. A characterization of perfect graphs. *Journal of Combinatorial Theory, Series B*, 13(2):95–98, 1972.

[52] L. Lovász. Normal hypergraphs and the perfect graph conjecture. *Discrete Mathematics*, 2(3):253–267, 1972.

[53] L. Lovász. Perfect graphs. *Selected topics in graph theory*, 2:55–87, 1983.

[54] L. Lovász, M. Grötschel, and A. Schrijver. Geometric algorithms and combinatorial optimization. *Berlin: Springer-Verlag*, 33:34, 1988.

[55] H. Meyniel. A new property of critical imperfect graphs and some consequences. *European Journal of Combinatorics*, 8(3):316, 1987.

[56] J. Nieminen. On the structure of minimum communication overlap networks. *IEEE transactions on circuits and systems*, 35(11):1465–1467, 1988.

[57] J. Nieminen and M. Peltola. The minumum geodetic communication overlap graphs. *Applied mathematics letters*, 11(2):99–102, 1998.

[58] R. Nussbaum and A-H. Esfahanian. Preliminary results on distance-preserving graphs. *Congressus Numerantium*, 211:141–149, 2012.

[59] R. Nussbaum, A-H. Esfahanian, and P-N. Tan. Clustering Social Networks Using Distance-Preserving Subgraphs. In *2010 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 380–385. IEEE, 2010.

[60] M.W. Padberg. Perfect zero–one matrices. *Mathematical Programming*, 6(1):180–196, 1974.

[61] P.A. Pevzner, H. Tang, and M.S. Waterman. An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 98(17):9748, 2001.

[62] J. Plesník. Towards minimum k-geodetically connected graphs. *Networks*, 41(2):73–82, 2003.

[63] J. Plesník. Minimum k-geodetically connected digraphs. *Networks*, 44(4):243–253, 2004.

[64] D.K. Pradhan and S.M. Reddy. A fault-tolerant communication architecture for distributed systems. *Computers, IEEE Transactions on*, 100(9):863–870, 1982.

[65] D. Ross, B. Sagan, R. Nussbaum, and A-H. Esfahanian. On constructing regular distance-preserving graphs. *Congressus Numerantium*, to appear.

[66] F. Roussel, I. Rusu, and H. Thuillier. The Strong Perfect Graph Conjecture: 40 years of attempts, and its resolution. *Discrete Mathematics*, 309(20):6092–6113, 2009.

[67] M.R. Samatham and D.K. Pradhan. The de Bruijn multiprocessor network: a versatile parallel processing and sorting network for VLSI. *IEEE Transactions on Computers*, pages 567–581, 1989.

[68] A. Sebö. The connectivity of minimal imperfect graphs. *Journal of Graph Theory*, 23(1):77–85, 1996.

[69] H. Whitney. Congruent graphs and the connectivity of graphs. In *Hassler Whitney Collected Papers*, pages 61–79. Springer, 1992.

[70] E. Zahedi. Distance-Preserving Graphs. *Congressus Numerantium*, to appear.

[71] D.R. Zerbino and E. Birney. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18(5):821, 2008.