WELL-DISTRIBUTED SETS ON GRAPHS

by

Jordan Barrett

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2018

Table of Contents

List of Tables

List of Figures

Abstract

Location theory is a topic widely researched in mathematics and computer science. The goal of this thesis will be to propose a new method for choosing vertices on a graph "optimally", in terms of spread, by generalizing concepts from music theory using physical interpretations. The sets from music theory are call *maximally even* and they have nice properties that one would expect to have when dealing with sets that are spread apart. However, these sets are only defined for directed cycles, and hence we must find a way to generalize the definition of *maximally even*. We introduce *well-distributed* sets as sets of charged particles repelling one another on a graph. We first show that this is indeed an extension of maximally even, after which we analyse well-distributed sets and classify them completely for some special families of graphs.

Acknowledgements

I would first like to thank my supervisor Dr. Jason Brown for collaborating with me and teaching me so much about graph theory and related concepts over the past year. Without Dr. Brown's mathematical experience and wisdom I would not have been able to write over half of this thesis. Having a supervisor who is so driven and enthusiastic about research has given me hope that I can achieve the same love for research as I progress through academia.

Next I would like to thank my friend and colleague Rebecca Ryan. Rebecca is the queen of organizing thoughts, formalizing concepts, and keeping me from pulling my hair out when I am stuck on a proof. To have such a close friend in the same line of research throughout my master's degree has made the whole process ten times more enjoyable and less stressful.

Lastly, I would like to thank Dr. Robert Dawson for explaining the behaviour of electromagnetic systems to me.

Chapter 1

Introduction

1.1 Location Theory

Location theory is a branch of computational geometry that, at its core, deals with the following question: given a graph G and a positive integer k, what is the optimal way to choose k vertices of G with some goal in mind and possibly under some restrictions? One example of this is the classic *facility location problem* [9]. In this formulation you must choose the location of k facilities (vertices) in G which minimizes the sum of the distances of each other vertex to its nearest facility. This type of problem might be looked at when deciding where to build hospitals, schools, fire departments, etc., in a large city. Another proposed solution is *dominating sets* [1]. A dominating set is a set of vertices in a graph such that, given any other vertex on the graph, it is adjacent to a vertex in the set. We will look at these two examples more closely. (In general, we follow "Graph Theory and Its Appliations" by Jonathan L. Gross and Jay Yellen for graph terminology [8].)

Definition 1.1. Let G be a graph and $S \subseteq V(G)$. For any vertex $v \in V(G)$, define the *distance from v to S*:

$$\text{dist}_G(v, S) = \min\{\text{dist}_G(v, u) : u \in S\}$$

Note that $\text{dist}_G(v, S) = 0$ if and only if $v \in S$.

Remark 1.1. This definition allows for both directed and undirected edges. In practice, $\text{dist}_G(v, S)$ is the fastest route *from v to S*.

Definition 1.2. Let G be a graph and $S \subseteq V(G)$. The *total distance from $V(G)$ to S* is defined as:

$$\sum_{v \in V(G)} \text{dist}_G(v, S)$$

We can now formalize the facilities location problem.

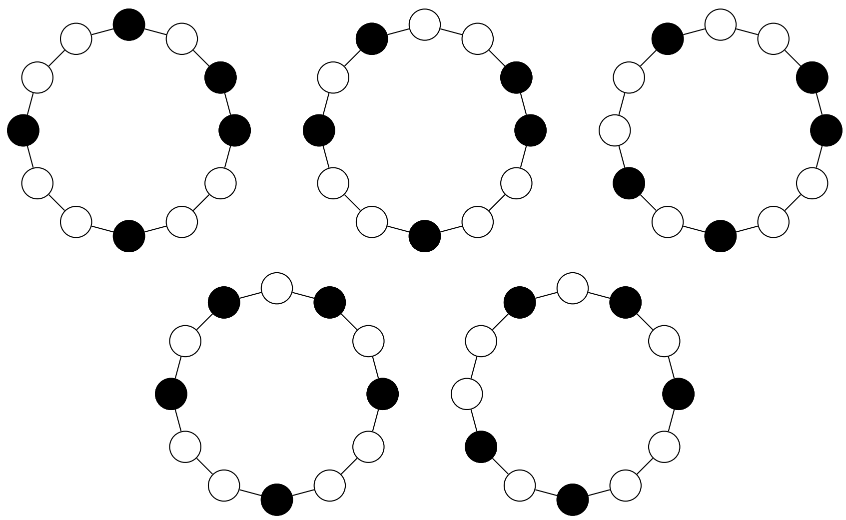


Figure 1.1: All facilities location solutions of size 5 in C_{12}

Definition 1.3. Let G be a graph of order n and let $k \in \{0, 1, \ldots, n\}$. We say the vertex set $S \in V(G)$ is a *facilities location solution* in G if:

1. $|S| = k$, and
2. For all $S' \in V(G)$, if $|S'| = k$ then the total distance from $V(G)$ to S' is at least the total distance from $V(G)$ to S.

In other words, S is a facilities location solution if it minimizes the sum of distances from vertices to S among all sets of the same size. Facilities location solutions sometimes seem impractical. Many solutions yield facilities that would never be reached, and facilities that would become crowded by the number of vertices that are closest to it.

Example 1.1. Let $G = C_{12}$, the undirected cyclic graph on 12 vertices, and let $k = 5$. A 5-facilities location solution in G would be any set S of size 5 where the total distance from $V(G)$ to S is 7, since there will be 7 unselected vertices (whose distance to S must be at least 1) and it is possible for each to be of distance 1 away from S. Figure 1.1 shows all possible 5-facilities location solutions in C_{12}, up to rotational and mirror symmetry, with the chosen sets coloured black: although these are all solutions, the second row of sets *look* more evenly spaced apart than the first, due to the sets in the first row each having adjacent vertices. However, from a facilities location point of view, all solutions are equally good. Hence, facilities locations might not necessarily be the *best* way to define evenly spaced sets.

Next we will formally define a dominating set, using the notation $N_G[S]$ to be the closed neighbourhood of a subset S of vertices in G:

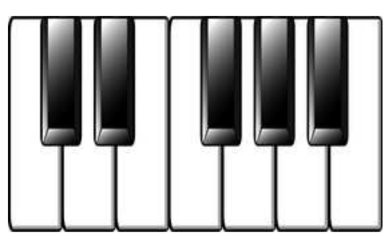
Definition 1.4. A set $S \subseteq V(G)$ is a dominating set if $N_G[S] = V(G)$.

Example 1.2. All of the sets in Figure 1.1 are also dominating sets. Thus, for the same reason as mentioned before with facilities location sets, dominating sets might not be the *best* way to define even spacing.

The question that arises from these examples is: given G and n, what is the *best* way to choose k vertices of G in such a way that the vertices are *spread apart* as much as possible? Obviously there is some subjectivity to this question. Certainly the solutions to a facilities location problem could be perceived as the *best*. However, the previous example would lead us to believe that there is a stronger condition that needs to be met.

1.2 A Concept from Music Theory

We seemingly digress now and talk about scales in music theory (but for good reason, as we shall see). The keys of a piano form a familiar pattern to musicians. The twelve fundamental notes that make up the chromatic scale are laid out on the piano like so, starting at middle C:



From left to right these notes are: $C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B$. The fundamental pitch frequencies which give rise to each note (given in Hz, i.e. cycles per second), rounded to the nearest integer, are: C : 262, $C\#$: 277, D : 294, $D\#$:

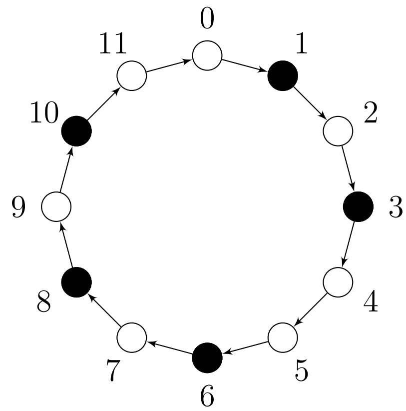


Figure 1.2: A cyclic representation of a piano

$311, E : 330, F : 349, F\# : 370, G : 392, G\# : 415, A : 440, A\# : 466, B : 494$. The rest of the frequencies on a piano (or any instrument for that matter) are related to one of the fundamental frequencies by some power of 2. In music theory, multiplying a frequency by 2 is the same as playing the note an octave above, which to most people sounds "the same". If a frequency is 2^n times another frequency ($n \in \mathbb{Z}$), the notes are n octaves apart. This relation is an equivalence relation, and so every octave of a pitch receives the same letter representation. One could represent the chromatic scale as $\mathbb{Z}_{12} = \mathbb{Z}/12\mathbb{Z}$ by mapping the letters $C, C\#, \ldots, B$ to the values $0, 1, \ldots, 11$ respectively. Then, instead of powers of 2, pitches would be related to one of the 12 fundamental frequencies via adding or subtracting some multiple of 12. If we treat the chromatic scale this way, we could represent the piano as the directed cyclic graph of order 12, which we will denoted $C_{12}^{\rightarrow}$, as shown in Figure 1.2. Given this labelling, we can see that the white vertices represent the white keys on a piano, and the black vertices represent the black keys. Notice the similarity between this graph and the very last graph in Figure 1.1. They are the same (under a rotation).

Remark 1.2. In general, chromatic scales of any positive integer size can be constructed under certain rules and customs. A general *equal tempered* chromatic scale of size n, with a root frequency r Hz, is comprised of the following set of fundamental frequencies:

$$\left\{r \cdot 2^{k/n} : k \in \{0, 1, \ldots, n-1\}\right\}$$

Given a general chromatic scale, we could represent it as a directed cyclic graph of order n, just as we have done with the "usual" chromatic scale with 12 notes. (The reason the standard chromatic scale contains 12 notes is, in part, due to the good approximations of "small" fractions.

In music theory there is a notion of a *maximally even* set of tones in the chromatic scale. This term was first defined by John Clough and Jack Douthett in their paper "Maximally Even Sets" [3]. They reconstruct the chromatic scale just as we have above, and introduce some necessary terminology. The following definitions here are restatements of Clough and Douthett's definitions, with the goal of re-wording and adapting notation more attuned to graph theory. Some of these are familiar to mathematicians, and others to musicians. (Throughout this thesis, we will take $\mathbb{N}$ to be the positive integers $\{1, 2, \dots\}$.)

Definition 1.5. The *chromatic universe* of size $n \in \mathbb{N}$, denoted $\mathcal{U}_n$, is the set $\{0, 1, \dots, n-1\}$. Given such a universe, and given $a, b \in \mathbb{N} \cup \{0\}$ such that $0 \leq a, b \leq n-1$, the *chromatic distance* from a to b in $\mathcal{U}_n$ is the smallest non-negative integer congruent to $(b-a) \pmod{n}$.

To reiterate Remark 1.2, we can represent the chromatic universe $\mathcal{U}_n$ as $C_n^{\rightarrow}$. With this representation, chromatic distance is equivalent to graph distance.

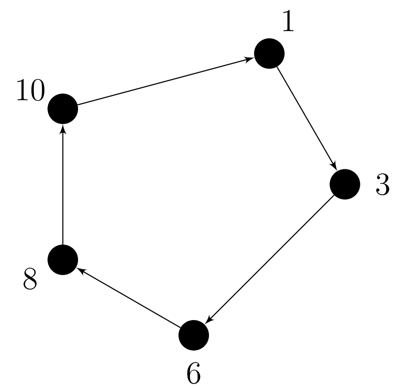
Definition 1.6. Given $n \in \mathbb{N}, k \in \mathbb{N} \cup \{0\}$ with $k \leq n$, a *scale* of size k in $\mathcal{U}_n$ is any subset of $\mathcal{U}_n$ of size k.

Continuing with our graphical interpretation, a *scale* is just a vertex subset. The set $S_k \subseteq V(C_n^{\rightarrow})$ will stand for a scale of size k. We will label and order S_k as the set $\{v_{s_1}, v_{s_2}, \dots, v_{s_{k-1}}\}$ with $s_0 < s_1 < \cdots < s_{k-1}$.

The next definition introduces a new distance that will be important for defining maximal evenness. From here on out we will use the graph theoretic notation.

Definition 1.7. Given $C_n^{\rightarrow}$, S_k, the *scale distance* from v_{s_i} to v_{s_j} is the smallest non-negative integer congruent to $(j-i) \pmod{k}$.

Example 1.3. Recall Figure 1.2. If we choose S_k as the set of black keys, we can think of the scale distance as the graph distance in the following graph:



In music terms, the chromatic distance between notes is the number of ascending half steps needed to get from the first note to the second. The scale distance would be the number of ascending note steps in the scale.

The next definition considers the relation between general distance and chromatic distance.

Definition 1.8. Given $C_n^{\rightarrow}$ and S_k, the *spectrum of chromatic distances* for a given scale distance d is a set of chromatic distances, denoted $< d >$, defined as:

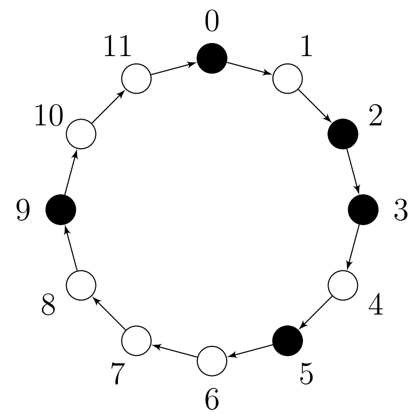
$$< d >= \{\text{dist}_{C_n}(v_i, v_j) : v_i, v_j \in S_k, \text{and the scale distance from } v_i \text{ to } v_j \text{ is } d\}$$

If the spectrum of chromatic distances for d is $d_1, d_2, \ldots, d_r$, we write $< d >= \{d_1, d_2, \ldots, d_k\}$. Note that trivially $< 0 >= \{0\}$, and we thus exclude it from any considerations.

Example 1.4. Recall Figure 1.2 once again with the scale set being the black keys (this set is a well known scale in music, called the *pentatonic* scale). The scale distances in this set are 1, 2, 3 and 4. The spectrum of chromatic distances for each scale distance is:

$$< 1 > = \{2, 3\}$$
$$< 2 > = \{4, 5\}$$
$$< 3 > = \{7, 8\}$$
$$< 4 > = \{9, 10\}$$

Example 1.5. Here is another set of 5 vertices:



The spectra of chromatic distances in this case are:

$$< 1 > = \{1, 2, 3, 4\}$$
$$< 2 > = \{3, 5, 6, 7\}$$
$$< 3 > = \{5, 6, 7, 9\}$$
$$< 4 > = \{8, 9, 10, 11\}$$

Notice the relation between the spectra and the *evenness* of these sets. In the first example the set appears to be spread apart as evenly as possible on the cycle. This gives rise to spectra each consisting of two consecutive positive integers. In the second example the set appears to be clumped together and not well spread apart, with the spectra each of size 4 and not necessarily with consecutive integers. This observation motivates the next definition.

Definition 1.9. Given $C_n^{\rightarrow}$ and S_k, we say S_k is *maximally even* in $C_n^{\rightarrow}$ if each spectrum of general distances corresponding to a specific distance is a single positive integer, or two consecutive positive integers.

The pentatonic scale (see Figure 1.2) is an example of a maximally even set, as we have shown in Example 1.4. The set in Example 1.5 was not maximally even. This is consistent with the intuition that the pentatonic scale is evenly spaced, while the clumps of vertices in Example 1.5 are not.

The following fundamental theorems all proven by Clough and Douthett [3].

Theorem 1.1 (Clough & Douthett)**.** *Given a chromatic universe $C_n^{\rightarrow}$ and $k \in \{0, 1, \ldots, n\}$, there exists a maximally even set of size k.*

Theorem 1.2 (Clough & Douthett)**.** *Given a chromatic universe $C_n^{\rightarrow}$, if $S_k = \{v_{s_0}, \ldots, v_{s_{k-1}}\}$ and $S'_k = \{v_{t_0}, \ldots, v_{t_{k-1}}\}$ are maximally even sets of the same size, then there exists $c \in \{0, 1, \ldots, n-1\}$ such that $v_{s_i} + c = v_{t_i}$ for all $0 \leq i \leq k-1$. That is, any two maximally even sets of size k are translates of one another.*

Theorem 1.3 (Clough & Douthett)**.** *Given a chromatic universe $C_n^{\rightarrow}$ and a maximally even set S_k, the complement $V(C_n^{\rightarrow}) \setminus S_k$ is also a maximally even set.*

These theorems tell us that for any directed cycle $C_n^{\rightarrow}$, there is a unique maximally even set (up to symmetry) of each size, and the complement of one such set is another such set.

Example 1.6. Figure 1.3 gives all the maximally even sets (up to symmetry) in $C_{12}^{\rightarrow}$. Many of these sets correspond to well known musical scales (or chords, if played simultaneously), and hence maximally even sets give rise to “good” sounding scales and chords.

Now we will circle back to graph theory. Choosing sets of vertices on graphs in such a way that they are spaced out as much as possible is an important practical tool in many areas of mathematics. We have explored the facilities location problem, as well as dominating sets, and found similar issues with both. The difference between these sets and maximally even sets is that maximal evenness is concerned with the spacing of vertices on a global scale. Every pair of vertices is taken into account when finding maximally even sets. On the other hand, facilities location solutions and dominating sets are only concerned with some local property, meaning vertices can still be “clumped together”, as seen in Example 1.1. This motivates us to try and generalize maximal evenness to arbitrary graphs.

1.3 Well-Distributed Sets in Graphs

We want to extend the notion of maximally even to arbitrary graphs. However, the concept of scale distance is not well defined on arbitrary graphs. Let G be a graph,

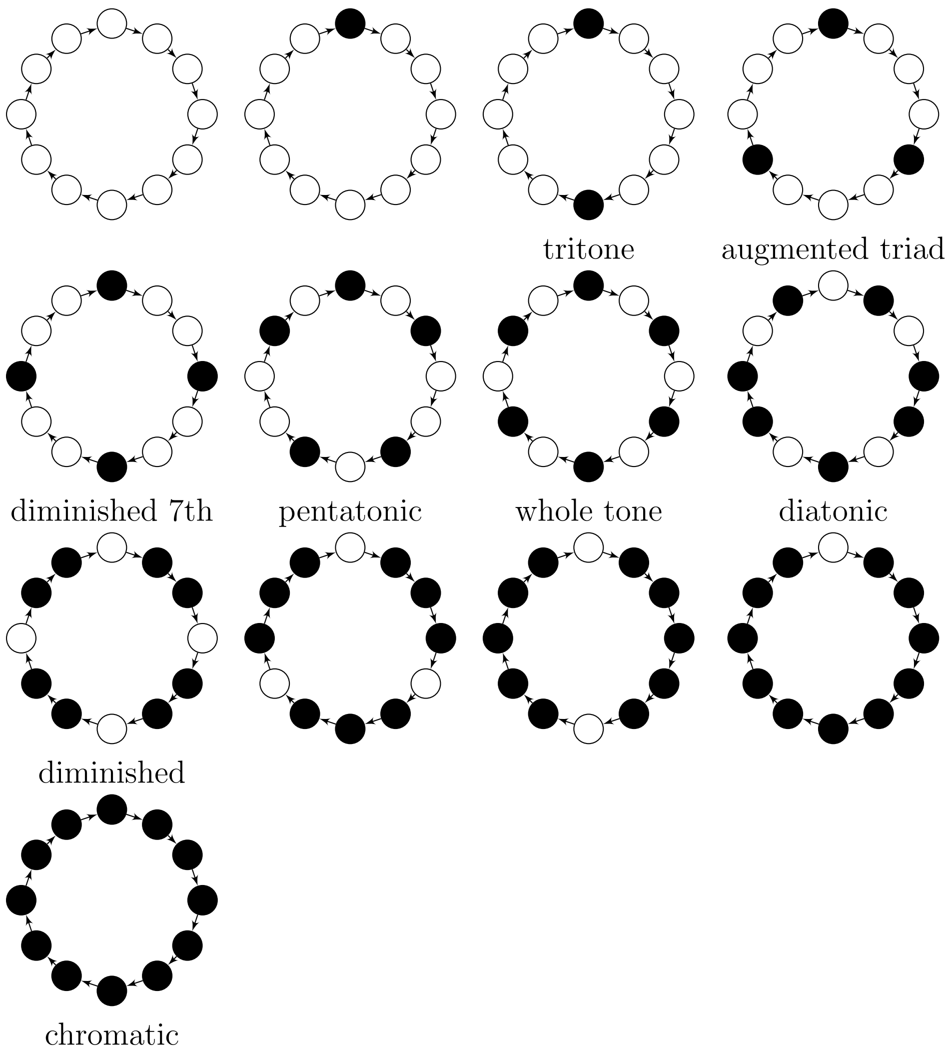


Figure 1.3: All maximally even sets in $C_{12}^{\rightarrow}$, up to a rotation.

$S \subseteq V(G)$, and $v_1, v_2 \in S$. There could be more than one shortest path from v_1 to v_2, and each path could contain a different number of vertices in S. If this is the case, what would be the analogous version of the scale distance from v_1 to v_2? Instead of trying to deal with this issue, we are going to try a different approach altogether, with a new motivation from physics.

Imagine a physical system in which some number of equally charged particles are placed in some space. The particles will repel away from each other until the system is stable. A stable state is one where, given any particle, the repelling forces from all other particles on the given particle have a net value of zero. One example of such a system is known as Thompson's problem in which one tries to understand how electrons would behave around the nucleus of an atom [11]. Thompson proposed the idea of a "plum pudding" model of the atom, in which electrons moved freely inside a sphere. The question became: how would a number of electrons spread apart inside a sphere under the inverse square law? Of course, this is no longer a relevant question in the structure of atoms, as newer models of the atoms have been pursued. However, the placement of equally charged particles in a bounded space is still of interest.

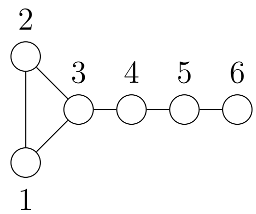
We can ask a similar question in a discrete space. Suppose we have a graph G, where we treat $V(G)$ as the potential positions for a set of particles. Then given such a set, how can we place it on the graph in such a way that the total energy of the system is minimized? Let us be precise as to what this means first:

Definition 1.10. Let G be a graph (directed or undirected) and let $S \subseteq V(G)$. The *total energy* of S in G is defined as:

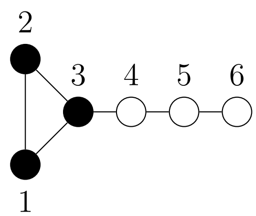
$$E(G, S) = \sum_{v_i, v_j \in S, i \neq j} \frac{1}{\operatorname{dist}_G(v_i, v_j)}. \tag{1.1}$$

This definition is not new by any means. The only differences between this and the total energy of an electromagnetic system are (a) the space is discrete, and dependent in G, (b) all of the masses/charges are treated as the same normalized values, and (c) for each pair v_i, v_j, we consider $\{v_1, v_2\}$ and $\{v_2, v_1\}$ as separate. We do this to allow one definition for both directed and undirected graphs.

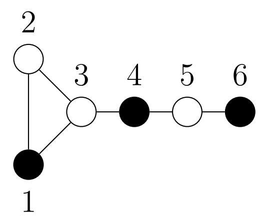
Example 1.7. Consider the following graph G:



Suppose we pick two different sets of vertices, $S_1 = \{1, 2, 3\}$ and $S_2 = \{1, 4, 6\}$.



S_1



S_2

It certainly *looks* like the vertices in S_2 are more spaced out than the vertices in S_1. Let us compare their energies.

$$\begin{aligned}
E(G, S_1) &= \frac{1}{\text{dist}_G(1,2)} + \frac{1}{\text{dist}_G(1,3)} + \frac{1}{\text{dist}_G(2,1)} \\
&\quad + \frac{1}{\text{dist}_G(2,3)} + \frac{1}{\text{dist}_G(3,1)} + \frac{1}{\text{dist}_G(3,2)} \\
&= \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} + \frac{1}{1} \\
&= 6
\end{aligned}$$

$$\begin{aligned}
E(G, S_2) &= \frac{1}{\text{dist}_G(1,4)} + \frac{1}{\text{dist}_G(1,6)} + \frac{1}{\text{dist}_G(4,1)} \\
&\quad + \frac{1}{\text{dist}_G(4,6)} + \frac{1}{\text{dist}_G(6,1)} + \frac{1}{\text{dist}_G(6,4)} \\
&= \frac{1}{2} + \frac{1}{4} + \frac{1}{2} + \frac{1}{2} + \frac{1}{4} + \frac{1}{2} \\
&= \frac{5}{2}
\end{aligned}$$

So S_2, the set that appears to be more spread apart, has a lower energy in this case.

Now that we have a better understanding of the total energy of a vertex set, we can define what it means for a subset of vertices to be *well-distributed* in a graph.

Definition 1.11. Given a graph G, the *minimum k-set energy* in G is defined as:

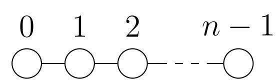
$$E(G,k) = \min\{E(G,S) : S \subseteq V(G), |S| = k\} \tag{1.2}$$

Definition 1.12. Let G be a graph and let $S \subseteq V(G)$ with $|S| = k$. We say S is *well-distributed* in G if $E(G,S) = E(G,k)$.

Remark 1.3. It turns out that, in Example 1.7, S_2 was in fact well-distributed.

We should look at some more examples to further justify this definition.

Example 1.8. Let P be an undirected path of length n:



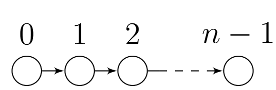
Clearly P has exactly one well-distributed set of size 2, i.e. $\{0, n-1\}$. Furthermore, any well-distributed set of size at least 2 must contain the set $\{0, n-1\}$: If we think of the set as charged particles, it makes sense that the leftmost particle would be pushed to the left end of the path, and similarly for the rightmost particle, as such a choice would drive down the energy.

As a quick proof, suppose the leftmost particle was in position $i > 0$. We could make a new set by replacing this particle with one in position 0. The total energy would only change by the pairs including the leftmost vertex, and those distances must all increase (and hence the reciprocals decrease). Thus, the leftmost particle must be in position 0. Similarly, the rightmost particle must be in position $n-1$.

What would be a well-distributed set of size 3? We need a j such that the set $\{0, j, n-1\}$ is minimized in P. Looking ahead, Lemma 2.2 will tell us that value must be $\frac{n-1}{2}$ if n is odd, or $\frac{n-1}{2} \pm \frac{1}{2}$ if n is even. Once $|S| > 3$ it becomes much more difficult to check if S is well-distributed.

Let us consider a similar example where we add directions to the edges.

Example 1.9. Let P be a directed path of length n:

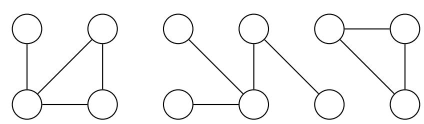


Let $S = \{i_1, i_2, \ldots, i_k\} \subseteq V(P)$ where $i_j < i_{j+1}$. Notice that $\text{dist}_P(i_a, i_b) = \infty$ whenever $a > b$, meaning $\frac{1}{\text{dist}_{P(i_a,i_b)}} = 0$. So if we let P_u be the undirected "version" of P, then $E(P,S) = \frac{1}{2}E(P_u, S)$. Consequently, S is well-distributed in P if and only if S is well-distributed in P_u.

Remark 1.4. The result from Example 1.9 only works because we look at *reciprocals* of distances instead of just distances themselves. If we had instead defined well-distributed sets as those that *maximize* the sum of all distances, any set of size at least 2 on P would have been well-distributed, since the sum of distances would be infinity.

Here is another example that justifies why we look at minimizing the sum of reciprocals of distances, rather than maximizing distances themselves:

Example 1.10. Consider a graph G with 3 components: G_1, G_2, and G_3 like so:



If we choose any set of size 3, with each vertex in a different component of G, the resulting total energy will be zero. Thus, any such set is well-distributed. This also implies that any set of size 3 with at least 2 vertices in the same component is not well-distributed, since the total energy will be greater than zero. However, if we took the sums of distances, both cases would yield ∞.

Example 1.10 gives rise to the following result:

Proposition 1.1. *Let G be a graph with k components and let $S \subseteq V(G)$ be well-distributed in G. Then for any component $G_i \subseteq G$, $S \cap V(G_i)$ is well-distributed in G_i. Furthermore:*

1. *If $|S| \leq k$ then $|S \cap V(G_i)| \leq 1$ for all components $G_i \subseteq G$.*

2. *If $|S| \geq k$ then $|S \cap V(G_i)| \geq 1$ for all components $G_i \subseteq G$.*

Proof. Let $S \subseteq V(G)$ be well-distributed in G and $G_i \subseteq G$ be a component of G. Suppose, to the contrary, that $S \cap V(G_i)$ is not well-distributed in G_i. Build S' from S by replacing the subset $S \cap V(G_i)$ with a subset of the same size that *is* well-distributed in G_i. It is clear that, if $G_1, G_2, \ldots, G_k$ are the components of G, then

$E(G,S) = \sum_{j=1}^{k} E(G_i, S \cap V(G_i))$ since $\frac{1}{\mathrm{dist}_{G(v_1,v_2)}} > 0$ only when v_1 and v_2 are in the same component. It follows that $E(G,S') < E(G,S)$, contradicting our choice of S. Therefore $S \cap V(G_i)$ is well-distributed in G_i.

Furthermore, if $|S| \leq k$ then $E(G,S) = 0$ precisely when all vertices in S are in different components. Thus, well-distributed sets are those with each vertex in a different component.

On the other hand, if $|S| \geq k$ and there is some component G_i such that $S \cap V(G_i) = \emptyset$, then by pigeonhole principle there must be another component G_j containing at least 2 vertices in S. Construct S' from S by removing a vertex in G_j and adding a vertex in G_i. Then $|S'| = |S|$ and $E(G,S') < E(G,S)$, contradicting S being well-distributed in G. □

There is another result hidden in Example 1.8. We mentioned how, for well-distributed sets on a path, the leftmost and rightmost vertices in the set must be the endpoints. In general, if a vertex in a well-distributed set is a cut vertex then there should be a vertex from the well-distributed set in each component of $G - v$.

Proposition 1.2. *Let G be a graph, $S \subseteq V(G)$ be well-distributed, and $v \in S$. Suppose v is a cut vertex of G and $G_1, G_2, \ldots G_\ell$ are the components of $G - v$. Then*

1. *$|S| > \ell$, and*
2. *$G_i \cap S \neq \emptyset$ for each $i \in \{1, 2, \ldots \ell\}$.*

Proof. Suppose $G, S, v, G_1, G_2, \ldots G_\ell$ are defined as in the proposition. Notice that if $|S| \leq \ell$ then there must be at least one i such that $G_i \cap S = \emptyset$. Thus, it is enough to show the second condition is true.

Suppose to the contrary that $G_i \cap S = \emptyset$ for some $i \in \{1, 2, \ldots \ell\}$. Let $u \in V(G_i)$. Consider the set S^* constructed from S by replacing v with u. Let $v_j \in S$, and G_j be the component of $G - v$ containing v_j. Since v is a cut vertex of G, any path from v_j to u *must* include v. Thus, $\mathrm{dist}_G(v_j, u) > \mathrm{dist}_G(v_j, v)$. Consequently, $E(G, S^*) < E(G, S)$, contradicting S being well-distributed. Therefore, $G_i \cap S \neq \emptyset$ for all $i \in \{1, 2, \ldots \ell\}$. □

This result can help us rule out sets when asking if they are well-distributed. For example, none of the sets in Figure 1.4 can be well-distributed.

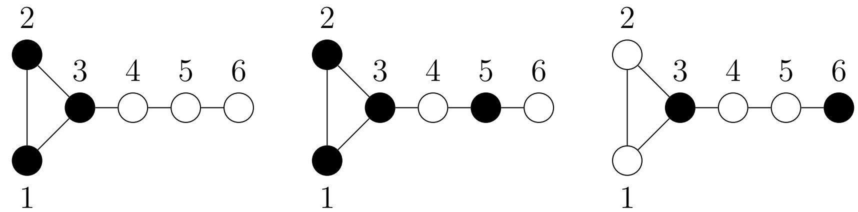


Figure 1.4: Sets that cannot be well-distributed due to Proposition 1.2

We have talked about maximally even sets defined on cyclic graphs. Then we introduced a new type of set called well-distributed. Both definitions give rise to sets which agree with our intuition of “spread apart”. However, it remains to show that well-distribution is indeed a generalization of maximally even.

Chapter 2

Equivalence of Definitions for Cycles

In the previous section we defined well-distributed sets on graphs and gave enough examples to justify the definition. What is important to show, however, is that well-distributed sets, when restricted to directed cycles, are precisely maximally even sets.

In fact, Douthett showed this to be true in his paper "The Theory of Maximally and Minimally Even Sets, the One-Dimensional Antiferromagnetic Ising Model, and the Continued Fraction Compromise of Musical Scales" [6]. This section will be dedicated to restating and reproving this theorem. The process, along with the intermediate results, will be similar, but with more rigour and with a graph theoretic approach. We will also generalize Douthett's result to *undirected* graphs, which are the graphs we are primarily interested in.

2.1 Directed Cycle Equivalence

We will state the theorem now then work towards proving it by the end of the chapter.

Theorem 2.1 (Douthett)**.** *Let C be a directed cycle and $S \subseteq V(C)$. S is maximally even in C if and only if S is well-distributed in C.*

Our plan for proving this result will be as follows:

1. Represent the well-distributed sets on a cycle as an optimization problem, where the function to be minimized is the total energy.

2. Create a new optimization problem from the original by relaxing the constraints, that is, by removing some of the restrictions.

3. Show that maximally even sets give rise to solutions of the generalized optimization problem.

4. Use this to show that maximally even sets are the precise solutions of the original optimization problem.

Remark 2.1. This process will be analogous to a common technique, called *LP-relaxation*, in solving integer programming problems. Suppose you are given an optimization problem with integer parameters. Make another optimization problem with the same optimizing value and restrictions but allow real solutions instead of integer ones. If you find integer solutions to the new problem, they must also be solutions to the original problem!

First, recall the notation from Section 1.2. Denote $C_n^{\rightarrow}$ as the directed cycle on n vertices with $V(C_n^{\rightarrow}) = \{v_0, v_1, \ldots, v_{n-1}\}$ so that there is an edge from v_i to v_{i+1} for all $0 \leq i < n-1$ and an edge from v_{n-1} to v_0.

Now let us define the optimization problem for a fixed n and k.

$$\min z = E(C_n^{\rightarrow}, S) \tag{2.1}$$

such that

$$\begin{array}{lcl}
S = \{v_{s_0}, v_{s_1}, \ldots, v_{s_{k-1}}\} \in V(C_n^{\rightarrow}) & & \\
s_0 < s_1 < \cdots < s_{k-1} & & \\
\text{dist}_{C_n^{\rightarrow}}(v_{s_0}, v_{s_1}) + \text{dist}_{C_n^{\rightarrow}}(v_{s_1}, v_{s_2}) + \cdots + \text{dist}_{C_n^{\rightarrow}}(v_{s_{k-1}}, v_{s_0}) & = & n \\
\text{dist}_{C_n^{\rightarrow}}(v_{s_0}, v_{s_2}) + \text{dist}_{C_n^{\rightarrow}}(v_{s_1}, v_{s_3}) + \cdots + \text{dist}_{C_n^{\rightarrow}}(v_{s_{k-1}}, v_{s_1}) & = & 2n \\
\vdots & \vdots & \vdots \\
\text{dist}_{C_n^{\rightarrow}}(v_{s_0}, v_{s_{k-1}}) + \text{dist}_{C_n^{\rightarrow}}(v_{s_1}, v_{s_0}) + \cdots + \text{dist}_{C_n^{\rightarrow}}(v_{s_{k-1}}, v_{s_{k-2}}) & = & (k-1)n.
\end{array}$$

We first need to show that these restrictions on the distances are what we want.

Lemma 2.1. *Given $C_n^{\rightarrow}$ and S_k, the following equality holds for all $1 \leq i \leq k-1$:*

$$dist_{C_n^{\rightarrow}}(v_{s_0}, v_{s_i}) + dist_{C_n^{\rightarrow}}(v_{s_1}, v_{s_{1+i}}) + \cdots + dist_{C_n^{\rightarrow}}(v_{s_{k-1}}, v_{s_{i-1}}) = i \cdot n$$

Proof. For the following, all subscripts are modulo k:

$$
\begin{aligned}
&\mathrm{dist}_{C_n^{\rightarrow}}(v_{s_0}, v_{s_i}) + \mathrm{dist}_{C_n^{\rightarrow}}(v_{s_1}, v_{s_{1+i}}) + \cdots + \mathrm{dist}_{C_n^{\rightarrow}}(v_{s_{k-1}}, v_{s_{i-1}}) \\
=&\sum_{j=0}^{i-1}\Big(\mathrm{dist}_{C_n^{\rightarrow}}(v_{s_j}, v_{s_{j+1}})\Big) + \sum_{j=1}^{i}\Big(\mathrm{dist}_{C_n^{\rightarrow}}(v_{s_j}, v_{s_{j+1}})\Big) + \cdots + \sum_{j=k-1}^{k-1+i-1}\Big(\mathrm{dist}_{C_n^{\rightarrow}}(v_{s_j}, v_{s_{j+1}})\Big) \\
=&\sum_{j=0}^{i-1}\Big(\mathrm{dist}_{C_n^{\rightarrow}}(v_{s_j}, v_{s_{j+1}}) + \mathrm{dist}_{C_n^{\rightarrow}}(v_{s_{j+1}}, v_{s_{j+2}}) + \cdots + \mathrm{dist}_{C_n^{\rightarrow}}(v_{s_{j+(k-1)}}, v_{s_{j+k}})\Big) \\
=&\sum_{j=0}^{i-1} n \\
=&i \cdot n
\end{aligned}
$$

□

Notice that (2.1) is just another way to represent the problem of finding well-distributed sets on cycles. Indeed, Lemma 2.1 shows that the equality conditions of (2.1) follow from the fact that S_k is a subset of vertices in $C_n^{\rightarrow}$ ordered in a certain way. So (2.1) could simply be written as follows:

$$\min z = E(C_n^{\rightarrow}, S)$$

such that

$$
\begin{aligned}
&S = \{v_{s_0}, v_{s_1}, \ldots, v_{s_{k-1}}\} \in V(C_n^{\rightarrow}) \\
&s_0 < s_1 < \cdots < s_{k-1}.
\end{aligned}
$$

The extra conditions are put there so that we can create a “generalized” optimization problem. Consider this new problem, for a fixed n and $k \leq n$:

$$\min z = \sum_{i,j=0, i\neq j}^{k-1} \frac{1}{d_{i,j}} \tag{2.2}$$

such that

$$
\begin{array}{lcl}
d_{0,1} + d_{1,2} + \cdots + d_{k-1,0} & = & n \\
d_{0,2} + d_{1,3} + \cdots + d_{k-1,1} & = & 2n \\
\vdots & \vdots & \vdots \\
d_{0,k-1} + d_{1,0} + \cdots + d_{k-1,k-2} & = & (k-1)n \\
d_{i,j} \in \mathbb{N}. & &
\end{array}
$$

What we mean by "generalized" is that $d_{i,j}$ does not necessarily have to represent a distance. However, it is clear by Lemma 2.1 that, for a given $C_n^{\rightarrow}$ and S_k, we could map $\mathrm{dist}_{C_n^{\rightarrow}}(v_{s_i}, v_{s_j})$ to $d_{i,j}$ and the resulting list $[d_{i,j}]_{i,j=0,i\neq j}^{k-1}$ would satisfy the *constraints* of (2.2). Thus, it makes sense to call (2.2) a generalization of (2.1).

Before continuing towards proving the main result, we need a lemma which helps explain the nature of the total energy equation:

Lemma 2.2. *Let $1 < a \leq b$, then for any $\alpha > 0$, $\frac{1}{a^\alpha} + \frac{1}{b^\alpha} < \frac{1}{(a-1)^\alpha} + \frac{1}{(b+1)^\alpha}$*

Proof. Let $1 < a \leq b$. Let $f(x) = x^{-\alpha} + (a+b-x)^{-\alpha}$ for $x \in (0, a+b)$. Then $f'(x) = \alpha\Big((a+b-x)^{-\alpha-1} - x^{-\alpha-1}\Big)$. So $f'(x) < 0$ if $x < \frac{a+b}{2}$ and $f'(x) > 0$ if $x > \frac{a+b}{2}$. Thus, f is a strictly decreasing function for $x \in \Big(0, \frac{a+b}{2}\Big)$. In particular, $a^{-\alpha} + b^{-\alpha} = f(a) < f(a-1) = (a-1)^{-\alpha} + (b+1)^{-\alpha}$, which proves the inequality. $\square$

Now we can prove the key lemma towards the theorem:

Lemma 2.3. *Let $D = [d_{j,j+i}]_{i=1,j=0}^{k-1,k-1}$ satisfy the constraints to (2.2). Then D is a solution set to (2.2) if and only if $|d_{j_1,j_1+i} - d_{j_2,j_2+i}| \leq 1$ for all $1 \leq i \leq k-1, 0 \leq j_1 < j_2 \leq k-1$.*

Proof. Suppose D is a solution set to (2.2). Now suppose, to the contrary, there was some i, j_1 and j_2 such that $|d_{j_1,j_1+i} - d_{j_2,j_2+i}| > 1$. Assume without loss of generality that $d_{j_1,j_1+i} < d_{j_2,j_2+i}$. Then by Lemma 2.2 we know that $\frac{1}{d_{j_1,j_1+i}+1} + \frac{1}{d_{j_2,j_2+i}-1} < \frac{1}{d_{j_1,j_1+i}} + \frac{1}{d_{j_2,j_2+i}}$. So define the new list $D^* = \big[d^*_{j,j+i}\big]_{i=1,j=0}^{k-1,k-1}$ by copying D and replacing the elements $d_{j_1,j_1+i}, d_{j_2,j_2+i}$ with the elements $d_{j_1,j_1+i}+1, d_{j_2,j_2+i}-1$ respectively. The resulting list D^* still satisfies the conditions of (2.2) since:

$$\begin{aligned}
&d^*_{0,i} + d^*_{1,1+i} + \cdots + d^*_{j_1,j_1+i} + \cdots + d^*_{j_2,j_2+i} + \cdots + d^*_{k-1,k-1+i} \\
&= d_{0,i} + d_{1,1+i} + \cdots + (d_{j_1,j_1+i} + 1) + \cdots + (d_{j_2,j_2+i} - 1) + \cdots + d_{k-1,k-1+i} \\
&= d_{0,i} + d_{1,1+i} + \cdots + d_{j_1,j_1+i} + \cdots + d_{j_2,j_2+i} + \cdots + d_{k-1,k-1+i} \\
&= i \cdot n
\end{aligned}$$

Furthermore:

$$\sum_{i,j=0,i\neq j}^{k-1} \frac{1}{d^*_{i,j}} < \sum_{i,j=0,i\neq j}^{k-1} \frac{1}{d_{i,j}}$$

contradicting the minimality of D.

Conversely, assume D satisfies the property $|d_{j_1,j_1+i} - d_{j_2,j_2+i}| \leq 1$ for all $1 \leq i \leq k-1, 0 \leq j_1 < j_2 \leq k-1$. Suppose, to the contrary, D was not a solution to (2.2). Then there must be a list D^* which *is* a solution. Then D^* also has the property that $\left|d^*_{j_1,j_1+i} - d^*_{j_2,j_2+i}\right| \leq 1$ for all $1 \leq i \leq k-1, 0 \leq j_1 < j_2 \leq k-1$. Thus, the lists $[d_{j,j+i}]_{j=0}^{k-1}$ and $\left[d^*_{j,j+i}\right]_{j=0}^{k-1}$ are list with either one element or two elements which are consecutive positive integers. Let q and $q+1$ be the two such values in the first list, and let q' and $q'+1$ be the two such values in the second list. Suppose in the first list there are r elements with value $q+1$ (r could be zero), and in the second list there are r' elements with value $q'+1$ (r' could be zero). What we get are the following two equations:

$$\begin{aligned} i \cdot n &= qk + r \qquad 0 \leq r < k \\ i \cdot n &= q'k + r' \qquad 0 \leq r' < k \end{aligned}$$

So $r \equiv i \cdot n \pmod{k}$, and r' is also this value. Therefore $r = r'$, and hence $q = q'$. Consequently, $[d_{j,j+i}]_{j=0}^{k-1}$ and $\left[d^*_{j,j+i}\right]_{j=0}^{k-1}$ are equivalent as multisets. Since i was arbitrary, $D \cong D^*$ up to the order of elements, contradicting the fact that D^* was a solution to (2.2) while D was not. □

With this result we can now prove the main theorem:

Proof of Theorem 2.1:
($\Rightarrow$) Let S_k be maximally even in $C_n^{\rightarrow}$ for some $0 \leq k \leq n$. Let us give values to the optimization problem (2.2) by setting $d_{j,j+i} = \text{dist}_{C_n^{\rightarrow}}(v_{s_j}, v_{s_{j+i}})$ for all $1 \leq i \leq k-1, 0 \leq j \leq k-1$. By Lemma 2.1 we know that the conditions of (2.2) are satisfied, and by Lemma 2.3 we know that this is actually a solution to (2.2). Consequently, $\left\{\text{dist}_{C_n^{\rightarrow}}(v_{s_j}, v_{s_{j+i}})\right\}_{i=1,j=0}^{k-1,k-1}$ is a solution set to (2.1). Thus, S_k is well-distributed in $C_n^{\rightarrow}$.

($\Leftarrow$) Let S_k be well-distributed in $C_n^{\rightarrow}$. Then $\left\{\text{dist}_{C_n^{\rightarrow}}(v_{s_j}, v_{s_{j+i}})\right\}_{i=1,j=0}^{k-1,k-1}$ is a solution set to (2.1). Recall that, in the context of maximal evenness, S_k is considered a scale in the chromatic universe C_n. Suppose for some $0 \leq a, b, c, d \leq k-1$, $v_{s_a}, v_{s_b}, v_{s_c}, v_{s_d} \in S_k$ were such that the scale distance from v_{s_a} to v_{s_b} equals the scale distance from v_{s_c} to v_{s_d}, and let this value be i. We have defined the vertices in S_k as $\{v_{s_0}, v_{s_1}, \ldots, v_{s_{k-1}}\}$

with $s_0 < s_1 < \cdots < s_{k-1}$. Thus, i is the smallest non-negative integer congruent to $b - a \pmod{k}$. Therefore, i is also congruent to $d - c \pmod{k}$. Thus, $b = a + i$ and $d = c+i$. It follows then by Lemma 2.3 that $\left|\text{dist}_{C_n^{\rightarrow}}(v_{s_a}, v_{s_{a+i}}) - \text{dist}_{C_n^{\rightarrow}}(v_{s_c}, v_{s_{c+i}})\right| \leq 1$. This means that, for any 2 pairs of vertices in S_k with the same scale distance, their chromatic distances differ by at most 1. This is precisely the definition of maximally even! $\square$

2.2 Undirected Cycle equivalence

To end the chapter we will generalize Theorem 2.1 to undirected cycles. This is important since only the directed case is found in [6], and we will be concerned only with undirected graphs from here on out.

Theorem 2.2. *Let C be an undirected cycle and $S \subseteq V(C)$. S is well-distributed in C if and only if S is well-distributed in $C^{\rightarrow}$, where $C^{\rightarrow}$ is obtained from C by directing C_n in one of the two circular orientations.*

Proof. Let C_n be an undirected cycle on n vertices, labeled analogously to the directed case. Start by defining a new optimization problem, where the right hand sides of the inequalities increase until the halfway mark, then decrease:

$$\min z = E(C_n, S) \tag{2.3}$$

such that

$$\begin{aligned}
&S = \{v_{s_0}, v_{s_1}, \ldots, v_{s_{k-1}}\} \in V(C_n) && \\
&s_0 < s_1 < \cdots < s_{k-1} && \\
&\text{dist}_{C_n}(v_{s_0}, v_{s_1}) + \text{dist}_{C_n}(v_{s_1}, v_{s_2}) + \cdots + \text{dist}_{C_n}(v_{s_{k-1}}, v_{s_0}) && \leq \; n \\
&\text{dist}_{C_n}(v_{s_0}, v_{s_2}) + \text{dist}_{C_n}(v_{s_1}, v_{s_3}) + \cdots + \text{dist}_{C_n}(v_{s_{k-1}}, v_{s_1}) && \leq \; 2n \\
&\vdots && \vdots \;\; \vdots \\
&\text{dist}_{C_n}(v_{s_0}, v_{s_{k-2}}) + \text{dist}_{C_n}(v_{s_1}, v_{s_{k-1}}) + \cdots + \text{dist}_{C_n}(v_{s_{k-1}}, v_{s_{k-3}}) && \leq \; 2n \\
&\text{dist}_{C_n}(v_{s_0}, v_{s_{k-1}}) + \text{dist}_{C_n}(v_{s_1}, v_{s_0}) + \cdots + \text{dist}_{C_n}(v_{s_{k-1}}, v_{s_{k-2}}) && \leq \; n.
\end{aligned}$$

This means there are two cases: one for even k and one for odd k. The even case is

$$\min z = E(C_n, S) \tag{2.4}$$

such that

$$\begin{array}{lcl}
S = \{v_{s_0}, v_{s_1}, \ldots, v_{s_{k-1}}\} \in V(C_n) & & \\
s_0 < s_1 < \cdots < s_{k-1} & & \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_1}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_2}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_0}) & \leq & n \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_2}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_3}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_1}) & \leq & 2n \\
\vdots & \vdots & \vdots \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_{\frac{k}{2}}}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_{\frac{k}{2}+1}}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_{\frac{k}{2}-1}}) & \leq & \frac{k}{2} \cdot n \\
\vdots & \vdots & \vdots \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_{k-2}}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_{k-1}}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_{k-3}}) & \leq & 2n \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_{k-1}}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_0}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_{k-2}}) & \leq & n.
\end{array}$$

The odd case is

$$\min z = E(C_n, S) \tag{2.5}$$

such that

$$\begin{array}{lcl}
S = \{v_{s_0}, v_{s_1}, \ldots, v_{s_{k-1}}\} \in V(C_n) & & \\
s_0 < s_1 < \cdots < s_{k-1} & & \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_1}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_2}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_0}) & \leq & n \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_2}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_3}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_1}) & \leq & 2n \\
\vdots & \vdots & \vdots \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_{\frac{k-1}{2}}}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_{\frac{k-1}{2}+1}}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_{\frac{k-1}{2}-1}}) & \leq & \frac{k-1}{2} \cdot n \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_{\frac{k+1}{2}}}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_{\frac{k+1}{2}+1}}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_{\frac{k+1}{2}-1}}) & \leq & \frac{k+1}{2} \cdot n \\
\vdots & \vdots & \vdots \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_{k-2}}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_{k-1}}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_{k-3}}) & \leq & 2n \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_{k-1}}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_0}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_{k-2}}) & \leq & n.
\end{array}$$

Notice that, in both cases, the i^{th} row and the $k-1-i^{th}$ row are identical since the distance function is symmetric for undirected graphs. So we can rewrite the problem as

$$\min z = E(C_n, S) \tag{2.6}$$

such that

$$\begin{array}{lcl}
S = \{v_{s_0}, v_{s_1}, \ldots, v_{s_{k-1}}\} \in V(C_n) & & \\
s_0 < s_1 < \cdots < s_{k-1} & & \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_1}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_2}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_0}) & \leq & n \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_2}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_3}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_1}) & \leq & 2n \\
\vdots & \vdots & \vdots \\
\mathrm{dist}_{C_n}(v_{s_0}, v_{s_{\lfloor \frac{k}{2} \rfloor}}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_{\lfloor \frac{k}{2} \rfloor + 1}}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_{\lfloor \frac{k}{2} \rfloor - 1}}) & \leq & \lfloor \frac{k}{2} \rfloor \cdot n.
\end{array}$$

What is important to notice is that the inequalities hold for any C_n and S_k. Indeed, since the following holds:

$$\mathrm{dist}_{(C_n)_\rightarrow}(v_{s_0}, v_{s_i}) + \mathrm{dist}_{(C_n)_\rightarrow}(v_{s_1}, v_{s_{1+i}}) + \cdots + \mathrm{dist}_{(C_n)_\rightarrow}(v_{s_{k-1}}, v_{s_{i-1}}) = i \cdot n$$

and since $\mathrm{dist}_{C_n}(u, w) \leq \mathrm{dist}_{(C_n)_\rightarrow}(u, w)$ for any $u, w \in V(C_n)$, we must have:

$$\mathrm{dist}_{C_n}(v_{s_0}, v_{s_1}) + \mathrm{dist}_{C_n}(v_{s_1}, v_{s_2}) + \cdots + \mathrm{dist}_{C_n}(v_{s_{k-1}}, v_{s_0}) \leq i \cdot n$$

Continuing the analogous argument, we will write general version of this optimization problem:

$$\min z = \sum_{i,j=0, i \neq j}^{k-1} \frac{1}{d_{i,j}} \tag{2.7}$$

such that

$$\begin{array}{lcl}
d_{0,1} + d_{1,2} + \cdots + d_{k-1,0} & \leq & n \\
d_{0,2} + d_{1,3} + \cdots + d_{k-1,1} & \leq & 2n \\
\vdots & \vdots & \vdots \\
d_{0,k-2} + d_{1,k-1} + \cdots + d_{k-1,k-3} & \leq & 2n \\
d_{0,k-1} + d_{1,0} + \cdots + d_{k-1,k-2} & \leq & n \\
d_{i,j} \in \mathbb{N} & &
\end{array}$$

The same logic applies as before, i.e. if we set $d_{j,j+i} = \mathrm{dist}_{C_n}(v_{s_j}, v_{s_{j+i}})$ and the result is a solution to (2.7), we must have started with a solution to (2.6). Furthermore, any solution to (2.7) must satisfy $d_{0,i} + d_{1,i+1} + \cdots + d_{k-1,i-1} = i \cdot n$ for all $0 < i < k$ since if $S = [d_{j,i+j}]_{j=0,i=1}^{k-1,k-1}$ were a solution such that $d_{0,i} + d_{1,i+1} + \cdots + d_{k-1,i-1} < i \cdot n$

then replacing $d_{0,i}$ with $d^*_{0,i} = d_{0,i} + 1$ would yield a strictly smaller z value and would still satisfy $d^*_{0,i} + d_{1,i+1} + \cdots + d_{k-1,i-1} \leq i \cdot n$, contradicting the minimality of S.

Since we can replace the inequalities by equalities, we can say by Lemma 2.3 that $D = [d_{j,j+i}]_{i=1,j=0}^{k-1,k-1}$ is a solution to (2.7) if and only if $|d_{j_1,i+j_1} - d_{j_2,i+j_2}| \leq 1$ for all i, j_1, j_2.

Note the following relation between C_n and $C_n^{\rightarrow}$:

$$\mathrm{dist}_{C_n^{\rightarrow}}(v_i, v_j) = \begin{cases} \mathrm{dist}_{C_n}(v_i, v_j) & \text{if} \quad j - i \leq n/2 \\ n - \mathrm{dist}_{C_n}(v_i, v_j) & \text{otherwise} \end{cases}$$

Consequently, if $|\mathrm{dist}_{C_n}(v_{s_{j_1}}, v_{s_{i+j_1}}) - \mathrm{dist}_{C_n}(v_{s_{j_2}}, v_{s_{i+j_2}})| \leq 1$ for all i, j_1, j_2 such that $i \leq n/2$, then $|\mathrm{dist}_{C_n}(v_{s_{j_1}}, v_{s_{i+j_1}}) - \mathrm{dist}_{C_n}(v_{s_{j_2}}, v_{s_{i+j_2}})| \leq 1$ must be true for all i, j_1, j_2. Hence, S is a solution if and only if S is maximally even. □

Thus, we are justified in generalizing maximal evenness via well-distribution. Now that this generalization has been established, we can now analyse the nature of well-distributed sets.

Chapter 3

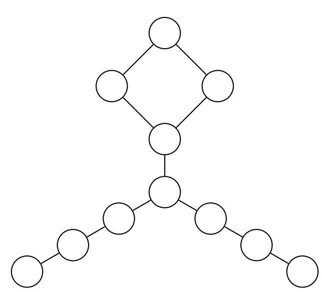
Analysing and Bounding Well-Distributed Sets

Now that we have a generalization of maximally even sets, we can begin to analyse the nature of these sets. From here on out we will assume all graphs are simple and undirected. We first want to know which results from maximally even sets carry over to well-distributed sets.

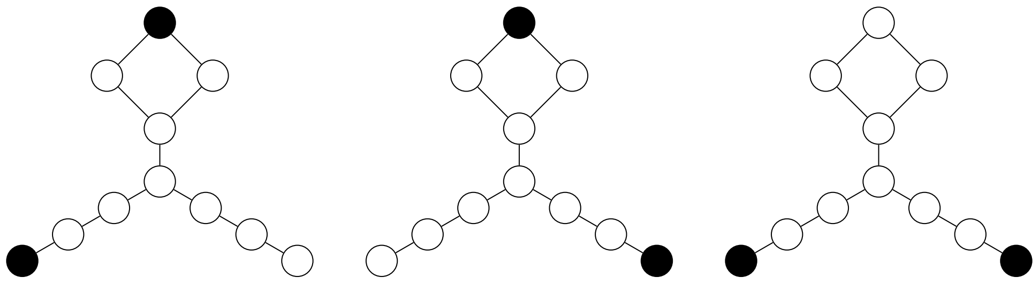
3.1 Extending Maximally Even Results to Well-Distributed Sets

Recall Theorems 1.1, 1.2 and 1.3, which collectively tell us that maximally even sets exists as unique, complementary pairs. Can we say something similar about well-distributed sets? Clearly well-distributed sets of a given size always exist by the way they are defined. However, is it always true that two well-distributed sets of the same size are "the same" under an automorphism.

Example 3.1. Let G be the following graph:



Looking ahead at chapter 5, we will learn that the well-distributed sets of size 2 in a connected graph are those pairs of vertices that achieve the diameter. Thus, there are three well-distributed sets of size 2. They are the following sets:



Notice that, while there is an automorphism of G that sends the first set to the second (and vice versa), there is no automorphism sending the third set to either the first or the second. Thus, Theorem 1.2 does not extend to well-distributed sets.

Just as there is no extension of Theorem 1.2, there is also no extension of Theorem 1.3. Example 1.8 showed that any well-distributed set of size at least 2 on a path must include both end points. Thus, the complement of a well-distributed set of size 2 in P_4 is not well-distributed.

Although these nice results of maximally even sets do not generalize to well-distributed sets, the intuitive property of being "spread apart" does. We will study these sets further to uncover more interesting properties.

3.2 Bounds on the Minimum Energy

What are some boundary conditions for the minimum k-set energy? Suppose a graph G has diameter d. Then any two vertices in G are at distance d or less from each other. Thus, we have a lower bound on the minimum energy.

$$E(G, k) \geq \frac{k(k-1)}{d}$$

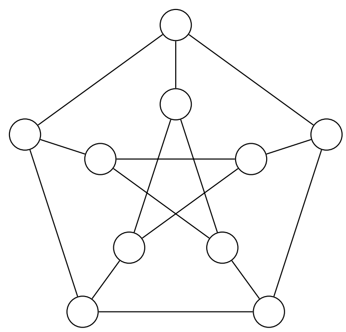
i.e. the extreme case where every pair of vertices is distance d apart. On the other hand, an upper bound for the minimum k-set energy would be the extreme case where every vertex is adjacent to every other vertex. In this case we have $E(G, k) = k(k-1)$. This upper bound would only be achieved on the complete graph, since any other graph would have at least one pair of vertices of distance at least two apart. However, this gives us a total bound on $E(G, k)$

Theorem 3.1. *Given a graph G with diameter d and a positive integer $k \leq V(G)$, the following bound holds:*

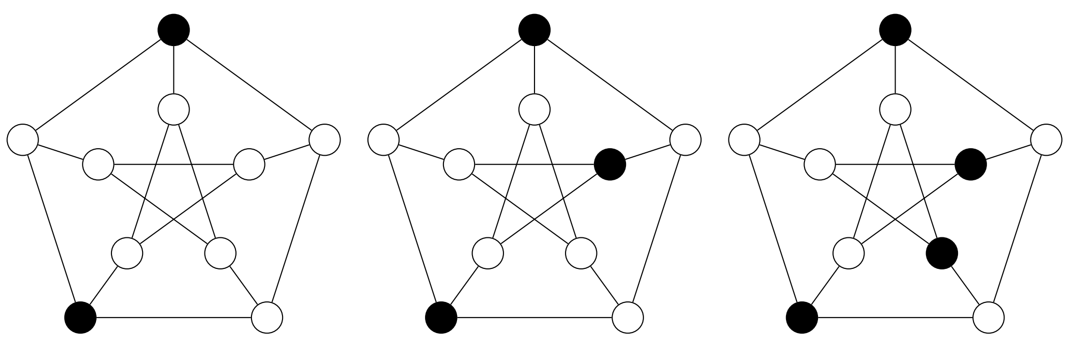
$$\frac{k(k-1)}{d} \leq E(G,k) \leq k(k-1)$$

These are rather extreme bounds. However, the lower bound serves as a sufficient condition for a set to be well-distributed.

Example 3.2. Let G be the Petersen graph, as shown below:



The diameter of this graph is 2, so the lower bound in Theorem 3.1 is useful. In particular, we can assure that the following are all well-distributed sets:



The lower bound is achieved in each of these cases, since all distances are 2. However, for any set of size greater than 4 in the Petersen graph, this bound cannot be achieved. To see this, notice that for any set of 3 inner vertices, 2 are adjacent. The same is true for 3 outer vertices.

We could look at other types of sets on graphs and use them to find boundary conditions. One such type would be independent sets [1].

Definition 3.1. Let G be a graph and $S \subseteq V(G)$. S is an *independent set* if no two vertices in S are adjacent.

A natural question that arises from independent sets is: given a graph G, what is the largest independent set? This is known as the *independence number* of a graph [1]. In other words:

Definition 3.2. Given a graph G, the *independence number* of G is the largest positive integer k such that there exists an independent set of size k in G.

Now suppose we have a graph G with independence number k and we are looking for a well-distributed set of size ℓ with $\ell \leq k$. Since we can find an independent set of size ℓ (take a subset of size ℓ of an independent set of size k), we can find a set of vertices with distances at least two from one another. Thus, the minimum ℓ-set energy is at most $\ell(\ell-1)/2$. If our set is larger than k, we can still find at least k points each with pairwise distance at least two. Thus, we have the following improved upper bound

Theorem 3.2. *Let G be a graph with independence number k. We have the following bounds for the minimum ℓ-set energy, depending on ℓ and k:*

- *If $\ell \leq k$ then*
$$E(G,\ell) \leq \frac{\ell(\ell-1)}{2}$$
- *If $\ell > k$ then*
$$E(G,\ell) \leq \ell(\ell-1) - \frac{k(k-1)}{2}$$

Proof. In either case, we can always find k points with pairwise distance at least two. Thus, the worse case scenario would be if each of those pairs were exactly distance two, and the rest of the vertices were all of distance one away from the rest of the set. Thus, in case 1 where $\ell \leq k$, any independent set S of size ℓ would satisfy the following equation:
$$E(G,S) \leq \frac{\ell(\ell-1)}{2}.$$
If $\ell > k$, let $S = S_1 \cup S_2$ where S_1 is an independent set of size k and S_2 is $\ell - k$ vertices chosen arbitrarily. Then the total energy of S_1 is at most $\frac{k(k-1)}{2}$. Combining this with the upper bound in Theorem 3.1, we get that
$$E(G,S) \leq \ell(\ell-1) - \frac{k(k-1)}{2}.$$
□

Example 3.3. Let G be constructed from K_{10} by removing three edges that form a triangle. Then for $\ell \in \mathbb{N}$ such that $\ell \leq 3$, $E(G, \ell) = \frac{\ell(\ell-1)}{2}$ and the well-distributed sets of size ℓ are the independent sets. For $\ell > 3$, we construct S with $|S| = \ell$ by choosing *the* independent set of order 3, plus any $\ell - 3$ other vertices. Since only three pairs of vertices in S are distance 2 apart, $E(G, \ell) = \ell(\ell-1) - \frac{\ell(\ell-1)}{2}$. Thus, the boundaries in Theorem 3.2 are achievable.

We can also find a lower bound on the minimum k-set energy via spanning subgraphs. Given a graph G and a spanning subgraph H, we know that $\text{dist}_G(v_i, v_j) \leq \text{dist}_H(v_i, v_j)$ for any pair of vertices in G, since H is obtained from G by (possibly) removing some of the edges in G. Thus, for a set of vertices S in G, $E(G, S) \geq E(H, S)$. This is particularly useful if we can find a spanning subgraph for which we know the well-distributed sets.

Example 3.4. Recall that a graph G on n vertices is *Hamiltonian* if there exists a subgraph H of G that is isomorphic to C_n [8]. In this case, the total energy of a well-distributed set of size k in G is at least the total energy of *the* maximally even set of size k in C_n.

3.3 Computational Complexity

We have talked about results coming from independent sets and graphs with diameter two. These results give us a way to show that the problem of finding well-distributed sets is an NP-complete problem. More specifically, finding the minimum k-set energy is an NP-complete problem.

WELL-DISTRIBUTED SET
INSTANCE: Undirected graph G, positive integer k and rational number r
QUESTION: Is $E(G, k) \leq r$

Clearly this problem is in NP since, if the answer is *yes*, one can simply give a set S that satisfies the inequality, and $E(G, S)$ can be calculated for any vertex set $S \subseteq V(G)$. Now we must show this is in fact NP-complete. We do this by reducing to the well known NP-complete problem of finding independent sets to it [7]. The problem in question is:

INDEPENDENT SET
INSTANCE: Undirected graph G, positive integer k
QUESTION: Does G have an independent set of cardinality k?

Given a graph G and a positive integer k, we are interested in whether G has an independent set of size k. We can assume $k \geq 2$, and hence C is not complete (if $k = 1$ then the answer is always yes!) Construct the graph $G+v$ by adding a vertex v to G and connecting v to every other vertex in G (this is known as adding a *universal vertex*). Observe that G has an independent set of size k if and only if $G+v$ does, since v cannot be included in the set, and no other edges have been added in the construction of $G+v$.

We can say the following about $G+v$:

1. $G+v$ can be constructed from G in polynomial time.

2. $G+v$ has diameter two.

We now show that $E(G+v,k) \leq k(k-1)/2$ if and only if G has an independent set of cardinality k. First, suppose $E(G+v,k) \leq k(k-1)/2$; as $\mathrm{diam}(G) = 2$, in any set S of cardinality k such that $E(G,S) \leq k(k-1)/2$, it must be the case where $\mathrm{dist}_G(u,v) = 2$ for all $u, v \in S$ (since the only possible distances are 1 and 2, and any set with a pair of adjacent vertices would yield a total energy larger than the lower bound of $k(k-1)/2$). Thus, S is an independent set of size k in $G+v$, meaning S is also an independent set of size k in G as $k \geq 2$.

Conversely, if G has an independent set S of size k, then $E(G+v,k) \leq k(k-1)/2$, as $E(G+v,k) \leq E(G+v,S) = \frac{k(k-1)}{2}$.

Since finding the well-distributed number is intractable, we will not search for an efficient algorithm to find them. However, we can still construct algorithms to find sets *nearly* well-distributed.

3.4 Approximation Algorithms

Consider once again the physical interpretation of well-distributed sets. We can think of the vertices in a set S as charged particles sitting in a graph G. Under this

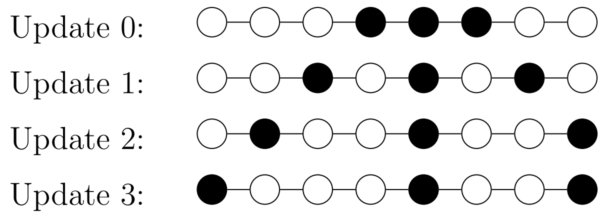


Figure 3.1: Output of the ‘spread’ algorithm on P_8 with a set of size 3

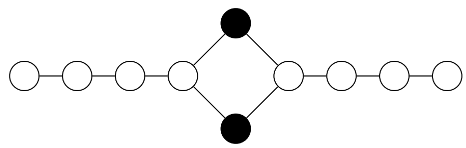
interpretation, the well-distributed sets are those that minimize the energy across *all* configurations. However, if one were to place charged particles randomly on a graph, they would not simply rearrange themselves into a well-distributed set. Instead they would move around the graph, continuously moving to adjacent vertices that yield smaller total energy. We can simulate this behaviour, as shown in the following pseudo code:

```
Procedure: Spread(G,S)
  while movement = true
    for v in S do
      make empty list
      for u in N(v) do
        add (u, E(G, S − v + u)) to list
      end for
      pick u in list with min energy and replace v with u
    end for
    print G
    if new S = old S then set movement = false
  end while.
```

Example 3.5. Figure 3.1 shows each step of the *spread* algorithm on P_8 with 3 particles. As expected, the leftmost and rightmost particles push to their respective ends at each iteration.

While the algorithm found a well-distributed set on a path, it will not necessarily find a well-distributed set in all cases.

Example 3.6. Consider the following graph and set of size 2:



Clearly this is not a well-distributed set. However, neither particle will move since it will initially increase the total energy.

Remark 3.1. Notice that we could extend the left and right paths in Example 3.6. This means that we cannot even bound the difference between $E(G, k)$ and $E(G, S)$ where S is the output of Spread(G, k).

Even though we cannot find well-distributed sets in this way, the algorithm will give us *something* along the lines of sets that are spread apart. Appendix A shows a number of such results. In each case, the left graph shows the initial configuration of points, and the right graph shows the final configuration. In some cases we are able to verify (either by exhaustive search or by previous knowledge) whether or not the final configuration is well-distributed. In other cases, we cannot verify due to limited computational power. This is indicated in the last column.

Even though we cannot find well-distributed sets in general, there are certainly special classes of graphs where we *can* classify them completely.

Chapter 4

Well-Distributed Sets on Special Families of Graphs

4.1 Complete and Complete Bipartite

Example 4.1. For a given $n \in \mathbb{N}$, the complete graph K_n and the edge-less graph O_n are trivial cases. Any set of vertices S in K_n of size k has $E(K_n, S) = k(k-1)$. Likewise, $E(O_n, S) = 0$.

Example 4.2. Given a complete bipartite graph $K_{m,n}$, there are only two possible distances between any two vertices. The distance is 1 if the vertices are on different independent sets, and 2 if they are on the same. Thus, the well-distributed sets of size less than or equal to $\max\{m, n\}$ are those that are in the same independent set. This is easily verifiable as any independent set satisfies the lower bound in Theorem 3.1, whereas any other set does not.

Remark 4.1. This argument can be generalized to show that the well-distributed sets of any size can be obtained by first filling up the larger of the two independent sets, then choosing the rest of the vertices arbitrarily from the smaller independent set. In fact, this can be further generalized to multipartite graphs. Given $K_{n_1,n_2,\ldots,n_k}$ with $n_1 \geq n_2 \geq \cdots \geq n_k$, well-distributed sets can be constructed by first filling up the n_1 independent set, then n_2, etc... We can think of this as trying to maximize the number of times we can add $1/2$ to the total energy and minimize the number of times we add 1, since they are the only two possibilities.

4.2 Random Graphs

For this section we follow "Random Graphs" by Béla Bollobás for probability and random graph terminology [2].

Can we say anything about well-distributed sets on random graphs? One of the most commonly used random graphs is the Erdös-Rényi $G(n, p)$ model [2].

Definition 4.1. Let n be a positive integer and $p \in (0,1)$. The random graph $G(n,p)$ is the set of all (simple, undirected) graphs on n vertices, and a probability distribution. The distribution is defined so that every pair of vertices is connected independently by an edge with probability p.

The random $G(n,p)$ graph has many nice properties. One such property is that, for a fixed $p \in (0,1)$, $\text{diam}(G(n,p)) = 2$ with probability tending to 1 (as $n \to \infty$). In fact, Bollobás proves in [2] that if we consider p as a function of n, then $\text{diam}(G(n,p)) = 2$ as $n \to \infty$ if $p^2 n - 2\log n \to \infty$ and $n^2(1-p) \to \infty$.

So we know that as n approaches infinity, the well-distributed sets in the $G(n,p)$ model are the independent sets, if such sets exist. We would therefore like to know the independence number of $G(n,p)$.

In [2], Bollabás shows that for fixed $p \in (0,1)$ and for fixed $\epsilon \in (0,1/2)$, the *clique number* of $G(n,p)$ tends to a value in the range of:

$$\Big((1+\epsilon)\log_{1/(1-p)} n, 2\log_{1/(1-p)} n\Big)$$

The *clique number* of a graph is the size of the largest complete subgraph. Since a complete subgraph in G corresponds to an independent set in the complement of G, we know that a complete subgraph in $G(n,p)$ corresponds to an independent set in $G(n,1-p)$, since the probability of G in $G(n,p)$ equals the probability of G^c in $G(n,1-p)$. Thus, the independence number of $G(n,p)$ tends to

$$C\log_{1/p} n$$

where $C \in (1+\epsilon, 2)$ is a constant.

Altogether this gives us the following Theorem:

Theorem 4.1. *For $p \in (0,1)$, $\epsilon \in (0,1/2)$, and $k \in \Big((1+\epsilon)\log_{1/p} n, 2\log_{1/p} n\Big)$, for any positive integer $\ell \leq k$, $E(G(n,p),\ell) = \ell(\ell-1)/2$ with probability tending to 1 as $n \to \infty$.*

Thus, we have a concrete result about the minimum k-set energy, under certain conditions, for the random $G(n,p)$ model.

In the next chapter, instead of classifying well-distributed sets in specific *graphs*, we will classify well-distributed sets of specific *sizes*.

Chapter 5

Small and Large Well-Distributed Sets

5.1 Well-Distributed Sets of Size Three

Well-distributed sets of size 3 are the smallest non-trivial sets to look at. To see this, first note that all sets of size 1 are well-distributed, and the sets of size 2 are precisely those which achieve the diameter (assuming the graph is connected; otherwise pick vertices from different components). We can already say quite a bit from previous results about these sets of size 3. In some cases we can explicitly say what the well-distributed sets are:

- For K_n and O_n, all sets of size 3.
- For $K_{m,n}$, if $\max\{m, n\} \geq 3$, all independent sets of size 3. Otherwise, fill the larger independent set first.
- For C_n, the maximally even set of size 3 up to symmetry.
- For P_n, the endpoints and the middle (or near middle if n is even).
- If G has two components, pick two on the larger diameter, and the third on the other component.
- If G has three or more components, pick any set with no two vertices in the same component.
- If $G = G(n, p)$, any independent set of size 3.

This is indeed a lot of information, but it is no where near a complete classification. Note that we can exhaustively search all sets of three vertices in a graph, which would give an answer in polynomial time. However, this does not tell us anything about the general classification. So we need to study these sets further.

Let us assume for the rest of the chapter that G is a connected, undirected graph, $S = \{v_1, v_2, v_3\}$ is a set of 3 vertices in G, and $d_{i,j}$ is the distance in G from v_i to v_j. We will fix an ordering on the distances, since v_1, v_2, v_3 can be permuted. So assume $d_{1,2} \leq d_{2,3} \leq d_{1,3}$.

Lemma 5.1. *A set S of cardinality 3 is well-distributed in G if and only if one of three conditions hold (up to symmetry of the vertices):*

1. *$d_{1,3} = diam(G)$ and v_2 lies halfway (or nearly halfway) along a diameter path from v_1 to v_3.*

2. *$d_{1,3} = diam(G)$ and $d_{1,2} + d_{2,3} > diam(G)$.*

3. *$\frac{diam(G)}{3} < d_{1,2} \leq d_{2,3} \leq d_{1,3} < diam(G)$*

Proof. First off, either $d_{1,3} = \mathrm{diam}(G)$ or $d_{1,3} < \mathrm{diam}(G)$.

Case 1: Assume $d_{1,3} = \mathrm{diam}(G)$. If v_2 is on a diameter path from v_1 to v_3, Lemma (2.2) tells us that $d_{1,2} = \lfloor \mathrm{diam}(G)/2 \rfloor$. So v_2 is halfway (or nearly halfway) between v_1 and v_3.

On the other hand, if v_2 is not on a diameter path from v_1 to v_3, we cannot have $d_{1,2} + d_{2,3} = d_{1,3}$. Nor can we have $d_{1,2} + d_{2,3} < d_{1,3}$ by the triangle inequality. Therefore, $d_{1,2} + d_{2,3} > d_{1,3} = \mathrm{diam}(G)$.

Case 2: Assume $d_{1,3} < \mathrm{diam}(G)$, so $\mathrm{diam}(G) > 1$. Suppose, to the contrary, $d_{1,2} \leq \mathrm{diam}(G)/3$. Since $d_{2,3}$ and $d_{1,3}$ are strictly less than $\mathrm{diam}(G)$ we have the following inequality:

$$\frac{1}{d_{1,2}} + \frac{1}{d_{2,3}} + \frac{1}{d_{1,3}} > \frac{3}{\mathrm{diam}(G)} + \frac{2}{\mathrm{diam}(G) - 1}$$

Now let us compare with the set $S' = \{u_1, u_2, u_3\}$ that satisfies the first condition in the lemma instead of the third. Then since $\mathrm{dist}_G(u_1, u_2)$ is at least $(\mathrm{diam}(G) - 1)/2$, and $\mathrm{dist}_G(u_2, u_3)$ is at least $\mathrm{diam}(G)/2$ we have:

$$
\begin{aligned}
\frac{1}{2}E(G,S') &= \frac{1}{\mathrm{dist}_G(u_1,u_2)} + \frac{1}{\mathrm{dist}_G(u_2,u_3)} + \frac{1}{\mathrm{dist}_G(u_1,u_3)} \\
&\leq \frac{2}{\mathrm{diam}(G)-1} + \frac{2}{\mathrm{diam}(G)} + \frac{1}{\mathrm{diam}(G)} \\
&< \frac{1}{d_{1,2}} + \frac{1}{d_{2,3}} + \frac{1}{d_{1,3}} \\
&= \frac{1}{2}E(G,S).
\end{aligned}
$$

This contradicts S being well-distributed. Thus, $\frac{\mathrm{diam}(G)}{3} < d_{1,2}$. □

Hidden in this lemma is an upper bound on the minimum 3-set energy of a connected graph G. Notice that, so long as $\mathrm{diam}(G) > 1$, we can always build a set that satisfied the first condition. Such a set would either be well-distributed, or there would be another set with less total energy. Thus, we can say the following:

Corollary 5.1.

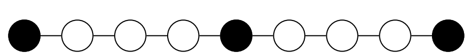
$$\frac{1}{2}E(G,3) \leq \frac{2}{diam(G)+1} + \frac{2}{diam(G)-1} + \frac{1}{diam(G)}.$$

Remark 5.1. If $\mathrm{diam}(G)$ is even we have a sightly better bound, since in this case we can pick v_2 exactly in the middle of v_1 and v_3, making $\mathrm{dist}_G(v_1,v_2) = \mathrm{diam}(G)/2$. Thus, the bound in this case is

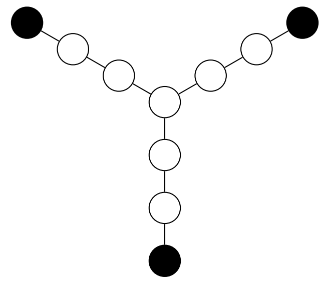
$$\frac{1}{2}E(G,3) \leq \frac{5}{\mathrm{diam}(G)}.$$

We have shown that these three cases are the only ones that *could* arise. We will give examples to show that each case is possible:

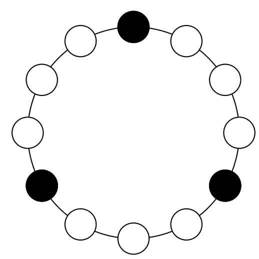
Example 5.1. Condition 1 is met for any path. For example, here is the well-distributed set of size 3 in P_9:



Condition 2 is met for any tree with three equal length branches. For example, here is the well-distributed set of size 3 on this star-type graph:



Condition 3 is met on cycles. Recall the maximally even set of size 3 in C_{12}, known in music theory as the *augmented triad*:

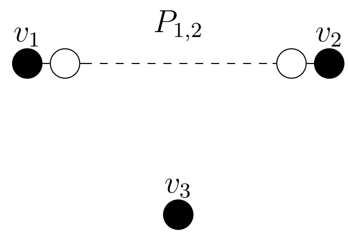


Lastly, we will classify all possible configurations of well-distributed sets of size three. Figure 5.1 shows some of the possible configurations that could arise. The next result shows that these subgraphs are in fact *all* possible subgraphs.

We will first prove a lemma to help exhaust all possible subgraphs.

Lemma 5.2. *Let G be a connected graph with at least 3 vertices, and let $v_1, v_2, v_3 \in G$. Let $P_{1,2} \subseteq V(G)$ be a shortest path from v_1 to v_2. Then, there exists a shortest path $P_{1,3}$ from v_1 to v_3 such that, if $P_{1,3}$ branches from $P_{1,2}$, then $P_{1,3}$ does not reconnect with $P_{1,2}$.*

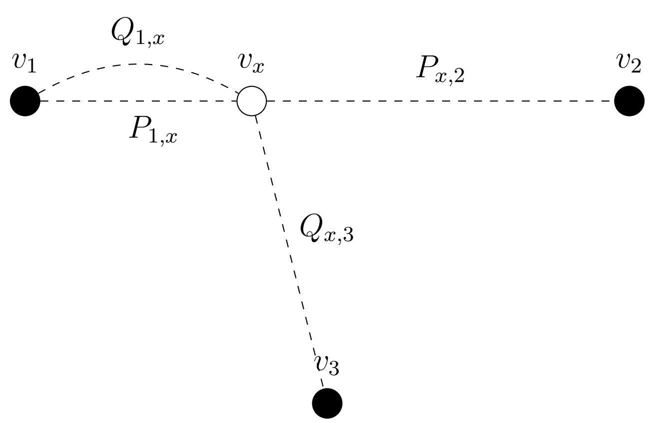
Proof. Suppose we have a shortest path $P_{1,2}$ from v_1 to v_2. First, if v_3 is on this shortest path, then a shortest path from v_1 to v_3 is simply the subset of $P_{1,2}$ from v_1 to v_3. Otherwise there would be a shorter path $P_{1,3}$ from v_1 to v_3; that is, $P_{1,3}$ plus the subset of $P_{1,2}$ from v_3 to v_2 would be a shorter path between v_1 and v_2 than $P_{1,2}$, contradicting $P_{1,2}$ being a shortest path. Therefore, we may assume $\{v_3\} \cap P_{1,2} = \emptyset$. Thus, we have the following set up:



Let us define a few more paths as follows:

- $Q_{1,3}$ is a shortest path from v_1 to v_3,
- v_x is the last vertex in $P_{1,2}$ that is touched when travelling from v_1 to v_3 along $Q_{1,3}$,
- $Q_{1,x}$ and $Q_{x,3}$ are the subpaths of $Q_{1,3}$ connecting v_1 to v_x and v_x to v_3 respectively, and
- $P_{1,x}$ and $P_{x,2}$ are the subpaths of $P_{1,2}$ connecting v_1 to v_x and v_x to v_2 respectively.

Visually, we have the following:



Since $P_{1,2}$ is a shortest path, $|P_{1,x}| \leq |Q_{1,x}|$ (otherwise $Q_{1,x} \cup P_{x,2}$ is a shorter path from v_1 to v_2). Likewise, since $Q_{1,3}$ is a shortest path, $|P_{1,x}| \geq |Q_{1,x}|$. Thus, $|P_{1,x}| = |Q_{1,x}|$. So let $P_{1,3} = P_{1,x} \cup Q_{x,3}$. Then $|P_{1,3}| = |Q_{1,3}|$, meaning $P_{1,3}$ is also a shortest path from v_1 to v_3. Furthermore, this path branches from $P_{1,2}$ at v_x, and does not reconnect to $P_{1,2}$ by the way we have defined v_x. □

Now we are ready to prove the main result.

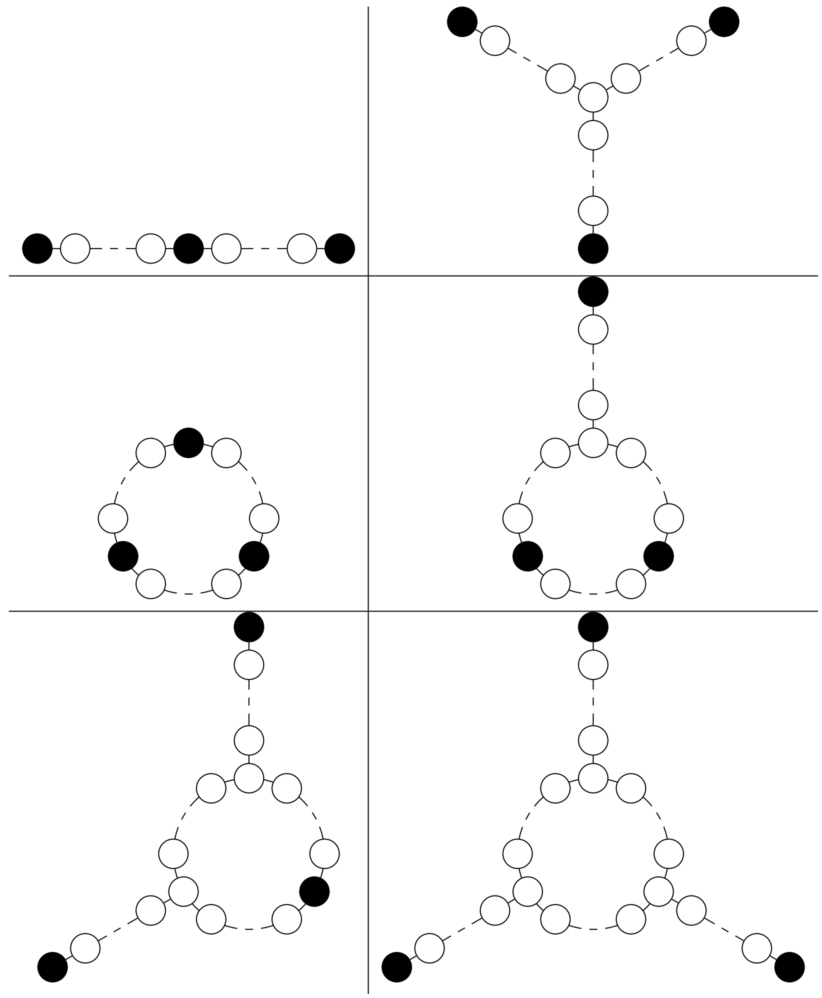
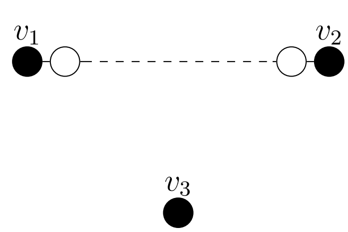


Figure 5.1: 6 possible subgraphs for well-distributed sets of size 3.

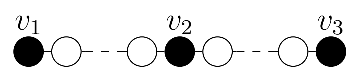
Proposition 5.1. *Given a graph G and a set S of three vertices in G, the generalized graphs in Figure 5.1 are all possible subgraphs of G containing S and a shortest path between each pair in S.*

Proof. We can start by picking a shortest path from any two vertices in G. Now we can exhaust all possible connections to the third vertex. Here is our set up:

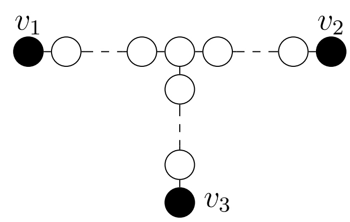


What are the possible shortest paths from v_1 to v_3? By Lemma 5.2, we have three possibilities:

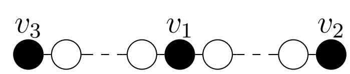
1. The paths could intersect at the whole path between v_1 and v_2. In this case we have the following subgraph:



2. The paths could intersect somewhere between v_1 and v_2. In this case we have the following subgraph:



3. The paths could intersect only at v_1. In this case we have the following subgraph:



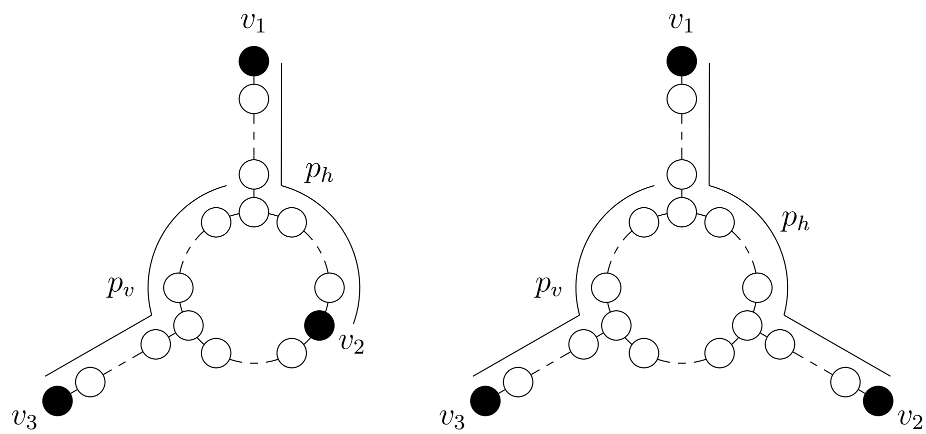
Now for each of the three cases we can find all possible connections between v_2 and v_3.

1. In this case, a shortest path between v_2 and v_3 must be the one already drawn, since it is part of a shortest path from v_1 to v_3.

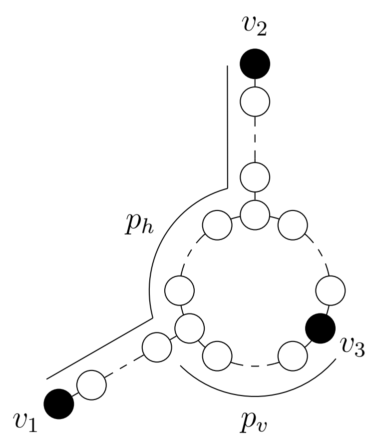
2. Let p_h be the path from v_1 to v_2 and p_v be the path from v_3 to p_h. Again, the shortest path from v_2 to v_3 will intersect both p_v and p_h in exactly one subpath. Furthermore, if the path intersected p_h left of the intersection of p_h and p_v, it would imply there was a shorter path from v_3 to v_1, which is a contradiction. With this information we can exhaust all possible cases:

   (a) The shortest path could include all of p_v. In this case, no new vertices are added and the subgraph is complete.

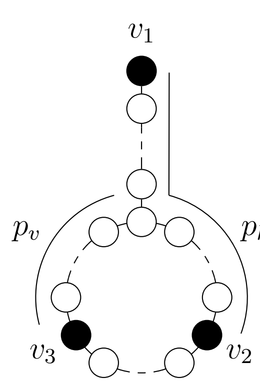
(b) The path could intersect at just v_2, or more than one vertex in p_h. The two subgraphs respectively are:



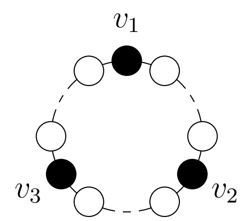
(c) The path could intersect just v_3 in p_v, and more than one vertex in p_h. This case is equivalent to a previous previous case, with a permutation of the vertices:



(d) The path could intersect at just v_2 and v_3. In this case we have:



3. In this last case, all outcomes are equivalent to outcomes in case 2 up to symmetry of v_1, v_2 and v_3, except for the case where the path intersects at just v_2 and v_3. This gives our last possibility:



$\square$

Thus, we have all possible configurations for well-distributed sets of size 3. For each case, we can determine where the vertices need to be in the subgraphs for the sets to actually be well-distributed. We will list the cases in the same order as Figure 5.1.

Case 1: In the case of the path we know that, with out loss of generality, $\text{dist}(v_1, v_3) = \text{diam}(G)$ and v_2 lies in the middle, or near middle, of v_1 and v_3.

Case 2: In the case of the star graph, each vertex must be as far away from the center as possible.

Case 3: In the case of the circle, the vertices must be the maximally even set of size 3 on the subgraph.

Case 4: In the case of the circle with one stick, let $C \subseteq G$ be the circle with circumference c, and $P \subseteq G$ be the stick with length p (where $C \cup P$ is the single-point intersection). Let v_1 and v_2 be the two vertices on the circle. Then by Lemma 2.2, we can deduce that $|\text{dist}_G(v_1, v_3) - \text{dist}_G(v_2, v_3)| \leq 1$.

Case 5: In the case of a circle with two sticks, we know by Lemma 2.2 that the distance from the vertex in the circle to the other two are equal, or differ by 1.

Case 6: In the case of a circle with three sticks, each vertex is a far away from the circle as possible.

This gives a complete classification of well-distributed sets of size 3. Next we will ask the opposite question. Namely: what are the well distributed sets of size close to $|V(G)|$.

5.2 Well-Distributed Sets on Nearly All Vertices

Given a graph G, there is only one well-distributed set of size $|V(G)|$, that set being $V(G)$. A more interesting question is: what are the well-distributed sets of size $|V(G)| - 1$?

Definition 5.1. Given a graph G, a set of vertices S, and $v \in S$, the *total in-energy* to v from S in G is

$$E_{in}(G, S, v) = \sum_{u \in S, u \neq v} \frac{1}{\mathrm{dist}_G(u, v)}.$$

The *total out-energy* from v to S is

$$E_{out}(G, S, v) = \sum_{u \in S, u \neq v} \frac{1}{\mathrm{dist}_G(v, u)}.$$

The *total energy* of v in S is

$$E(G, S, v) = E_{in}(G, S, v) + E_{out}(G, S, v)$$

Remark 5.2. If G is undirected then $E_{out}(G, S, v) = E_{in}(G, S, v) = 1/2E(G, S, v)$.

Now let $S = V(G)\backslash\{v\}$ for some vertex $v \in V(G)$. Then $E(G, S) = E(G, V(G)) - E(G, V(G), v)$. Thus, finding well-distributed sets of size $|V(G)| - 1$ is equivalent to finding $v \in V(G)$ with the *maximum* total energy in $V(G)$.

Example 5.2. Let $G = P_n$, with the vertices ordered from v_0 to v_{n-1}. Then the total energy of v_i in $V(P_n)$ is

$$E(P_n, V(G), v_i) = 2 \sum_{j=0, j\neq i}^{n-1} \frac{1}{|i - j|}$$

We can show that v_α where $\alpha = \lfloor (n-1)/2 \rfloor$ is a vertex with maximal total energy. Take some other positive integer i with $i < \alpha$. Then for some $c \in \{1, 2, \ldots, \alpha\}$, $j = \alpha - c$ (notice that this covers the first half of the path, and so by symmetry it covers all of the path). Thus, we have:

$$
\begin{aligned}
&\frac{1}{2}E(P_n, V(G), v_\alpha) - \frac{1}{2}E(P_n, V(G), v_{\alpha-c}) \\
&= \left(\sum_{j=0, j\neq\alpha}^{n-1} \frac{1}{|\alpha - j|} \right) - \left(\sum_{j=0, j\neq\alpha-c}^{n-1} \frac{1}{|\alpha - c - j|} \right) \\
&= \left(\sum_{j=1}^{\alpha} \frac{1}{j} + \sum_{j=1}^{n-1-\alpha} \frac{1}{j} \right) - \left(\sum_{j=1}^{\alpha-c} \frac{1}{j} + \sum_{j=1}^{n-1-\alpha+c} \frac{1}{j} \right) \\
&= \left(\sum_{j=1}^{\alpha} \frac{1}{j} - \sum_{j=1}^{\alpha-c} \frac{1}{j} \right) + \left(\sum_{j=1}^{n-1-\alpha} \frac{1}{j} - \sum_{j=1}^{n-1-\alpha+c} \frac{1}{j} \right) \\
&= \sum_{j=\alpha-c+1}^{\alpha} \frac{1}{j} - \sum_{j=n-\alpha}^{n-\alpha+(c-1)} \frac{1}{j} \\
&= \left(\frac{1}{\lfloor (n-1)/2 \rfloor - c + 1} + \cdots + \frac{1}{\lfloor (n-1)/2 \rfloor} \right) \\
&\quad - \left(\frac{1}{\lceil (n-1)/2 \rceil} + \cdots + \frac{1}{\lceil (n-1)/2 \rceil + c - 1} \right) \\
&= \sum_{j=0}^{c-1} \frac{1}{\lfloor (n-1)/2 \rfloor - (c-1) + j} - \frac{1}{\lceil (n-1)/2 \rceil + j} > 0
\end{aligned}
$$

So even in the simple case of the path, finding the well-distributed sets of size $n-1$ is non-trivial. However, we have shown that, when searching for well-distributed sets of this size, we need only look at the vertex *not* in the set. The same reasoning can be said about sets of size $n-2$ and $n-3$.

Chapter 6

Conclusion

The goal of this thesis was to propose a new family of sets, called well-distributed, on graphs that formally define the loose concept of being "spread apart" as much as possible and analyse such sets. These sets come from music theory, and are generalized via a physical interpretation. We have shown that well-distributed sets appeal to our intuition of being "spread apart" better than some other common sets in graph theory such as dominating sets and facilities location solutions. In particular, these sets never seem to "bunch up" in local groups. Since the energy function looks at *all* pairs of distances, we will always end up with sets that are *globally* spread apart in stead of *locally* spread apart.

There are many directions one could take to develop the theory of well-distributed sets. Here are some ideas for further research:

Generalizing the Energy Function

We can generalize the definition of well-distributed sets since we are not restricted to a physical system. Indeed, any reasonable energy function that is inversely proportional to *some power* of the distance should do the trick. Here is the generalized definition:

Definition 6.1. Let G be a graph and let $S \subseteq V(G)$ represent a set of equally charged particles in G. For any $\alpha > 0$, the *total α-energy* of S in G is defined as:

$$E(G, S, \alpha) = \sum_{v_i, v_j \in S, i \neq j} \frac{1}{\mathrm{dist}_G(v_i, v_j)^\alpha}$$

Setting $\alpha = 1$ is the physical choice. However, there are other choices for α that could prove beneficial. For example if $\alpha = 2$, the physical analogy would be to minimize the electromagnetic field, rather than to minimize the energy which is what the particles will want to do naturally.

Notice that the equivalence between well-distribution and maximal evenness on cycles did not require us to choose $\alpha = 1$. Indeed, the only time the energy function was needed was for Lemma 2.2, which is true for arbitrary $\alpha > 0$. Thus, we have the following generalized theorem:

Theorem 6.1. *Let C be a directed cycle, $S \subseteq V(C)$, and $\alpha > 0$. S is maximally even in C if and only if S is well-distributed with respect to α in C. In particular, S is well-distributed with respect to α in C if and only if S is well-distributed with respect to β in C, for all $\beta > 0$.*

This next example shows how changing α could change the nature of the well-distributed sets in certain situations.

Example 6.1. Let G be a graph with two components C_n and C_m, where C_n (C_m) is the directed cycle of length n (m). What are the well-distributed sets of size k? Proposition 1.1 and Theorem 6.1 tell us that if we choose a well-distributed set S_k with a vertices from $V(C_n)$ and $b = k - a$ vertices from $V(C_m)$, $S_k \cap C_n$ is the (up to symmetry) maximally even set of size a in C_n, and similarly for $S_k \cap C_m$. So suppose for the moment we chose a and b arbitrarily so that $0 \leq a, b \leq k$ and $a + b = k$. We can write down the exact equation for the total energy of this set.

First, consider C_n and S_a with the usual labeling. We know that for any $0 \leq i < a$, $\mathrm{dist}_{C_n}(v_{s_i}, v_{s_{i+1}}) = \left\lfloor \frac{n}{a} \right\rfloor$ or $\left\lfloor \frac{n}{a} \right\rfloor + 1$. In fact, we know exactly how many such distances are $\left\lfloor \frac{n}{a} \right\rfloor + 1$. There are r such distances, where r is the remainder of $n \pmod a$. In other words, $r = n - a \cdot \left\lfloor \frac{n}{a} \right\rfloor$. Similarly, $\mathrm{dist}_{C_n}(v_{s_i}, v_{s_{i+2}}) = \left\lfloor \frac{2n}{a} \right\rfloor$ or $\left\lfloor \frac{2n}{a} \right\rfloor + 1$, with $\left\lfloor \frac{2n}{a} \right\rfloor + 1$ occurring $2n - a \cdot \left\lfloor \frac{2n}{a} \right\rfloor$ times. Continuing with this, if we let r_i be the remainder of $i \cdot n \pmod k$, we end up with the following formula (we will look at an arbitrary α):

$$E(C_n, S_a, \alpha) = \sum_{i=1}^{a-1} \left(\frac{a - r_i}{\left\lfloor \frac{i \cdot n}{a} \right\rfloor^{\alpha}} + \frac{r_i}{\left(\left\lfloor \frac{i \cdot n}{a} \right\rfloor + 1 \right)^{\alpha}} \right)$$

With this, we can empirically find which proportion of a and b yields a well-distributed set for different values of α. For example, Table 6.1 shows the difference between $\alpha = 1$ and $\alpha = 2$ for $n = 100$ and $m = 50$. Notice that, in almost all cases, $a/b = n/m$ when $\alpha = 2$. However, when $\alpha = 1$, it seems $a/b < n/m$. This suggests that when $\alpha = 2$, the ratio of a to b is the same as the ratio of n to m, and that this is not the case for $\alpha = 1$.

$\|S\|$	$\|S_1\|$ ($\alpha = 1$)	$\|S_2\|$ ($\alpha = 1$)	$\|S_1\|$ ($\alpha = 2$)	$\|S_2\|$ ($\alpha = 2$)
1	1	0	1	0
2	1	1	1	1
3	2	1	2	1
4	2	2	3	1
5	3	2	3	2
6	4	2	4	2
7	4	3	5	2
8	5	3	5	3
9	6	3	6	3
10	6	4	7	3
20	13	7	13	7
30	19	11	20	10
40	25	15	26	14
50	33	17	33	17
75	50	25	50	25
100	64	36	67	33
125	79	46	83	42
150	100	50	100	50

Table 6.1: The distribution of vertices in well-distributed sets on $C_{100} \cup C_{50}$ for $\alpha = 1$ vs. $\alpha = 2$

The Well-Distributed Polynomial

Analogous to the *independence polynomial*, we can associate a *well-distributed polynomial* to each graph G (J. Brown, personal communication). Such a polynomial can associate combinatorial correlations between the minimum k-set number and analytic properties of the function.

Definition 6.2. Let G be a graph on n vertices. The *Well-distributed polynomial* on G is defined as:

$$P_{wd}(G, x) = \sum_{i=0}^{n} a_i x^i$$

where a_i is the number of well distributed sets of size i in G.

What are some of these polynomials? What are their roots?

Example 6.2. If G is either the complete graph or the empty graph, every set of vertices is well-distributed. Thus, the well-distributed polynomial of K_n (and also for O_n) is:

$$P_{wd}(K_n, x) = \sum_{i=0}^{n} \binom{n}{i} x^i = (x+1)^n$$

Thus, the only root of $P_{wd}(K_n, x)$ ($P_{wd}(O_n, x)$) is -1.

Example 6.3. If G has independence number k and $\operatorname{diam}(G) = 2$, then the polynomial

$$P_{wd}(G, x) - \sum_{i=k+1}^{n} a_i x^i$$

is the independence polynomial of G. This is because the well-distributed sets in G of size at most k are precisely the independent sets.

Example 6.4. Given C_n, the number of maximally even sets of size k in n is $n/\gcd(n, k)$ [6]. This gives the following polynomial:

$$P_{wd}(C_n, x) = \sum_{i=0}^{n} \frac{n}{\gcd(n, i)} x^i$$

This polynomial is similar to a class of polynomials studied by Karl Dilcher and Sinai Robins in their paper "Zeros and Irreducibility of Polynomials with GCD Powers as Coefficients" [4]. Dilcher and Robins prove that polynomials of the form

$$P(n, k) = \sum_{j=0}^{n} \gcd(n, j)^k z^j$$

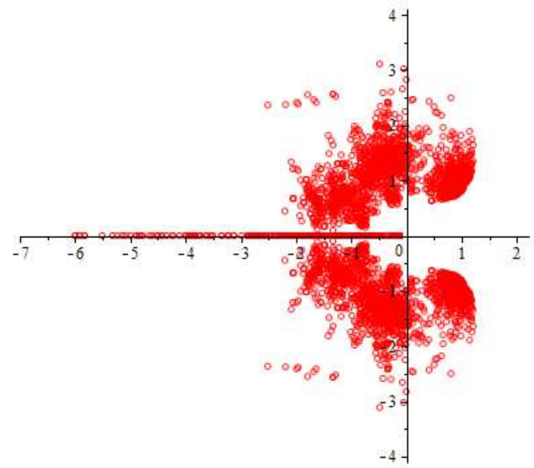


Figure 6.1: All elements of S_{wd} coming from graphs of order at most 7.

for $k \geq 1$, have all of their roots lying on the unit circle. However, the same cannot be said for $P_{wd}(C_n, x)$. For instance, $P_{wd}(C_4, x) = x^4 + 4x^3 + 2x^2 + 4x + 1$, which has roots at $x = -2 \pm \sqrt{3}$.

We would be interested in understanding the set of all possible roots of well-distributed polynomials. Let

$$S_{wd} = \{x \in \mathbb{C} : P_{wd}(G, x) = 0 \text{ for some graph } G\}.$$

We can make some conclusions with our analysis thus far. First off we know that if $x \in S_{wd}$ is real then $x < 0$, as x cannot be positive if all of the coefficients of well-distributed polynomials are positive, and x cannot be zero since $a_0 = 1$ for all graphs. Secondly, since $a_0 = a_n = 1$ for all graphs, we know by the rational root theorem that the only rational number in S_{wd} is -1.

We can find all such roots for small graphs using Maple. Figure 6.1 gives the set of roots of well-distributed polynomials for all connected graphs of order at most 7. Looking at these roots, we could make the following conjectures:

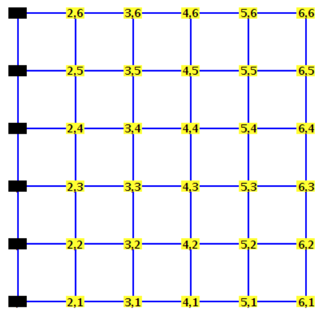
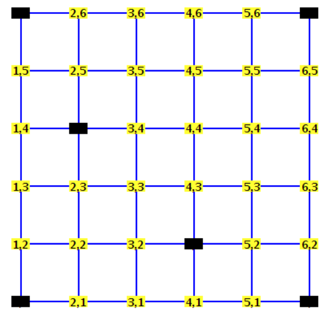
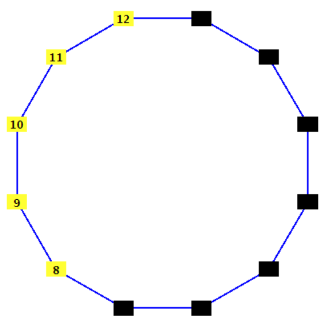
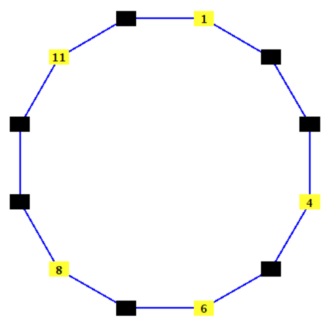
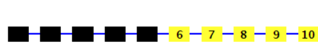
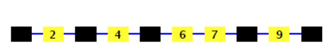
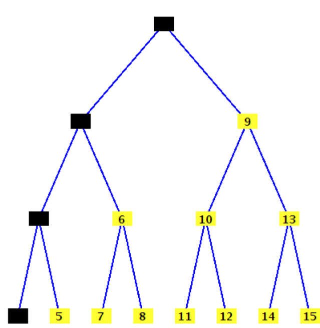
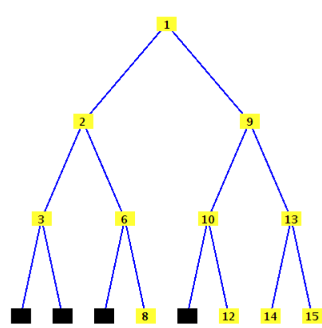
1. The set $\{x \in S_{wd} : |x| = 1\}$ is dense in the unit circle, and

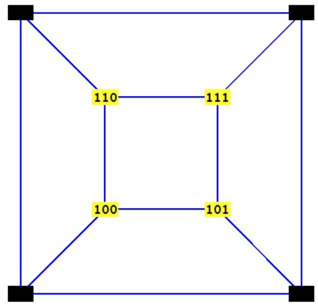
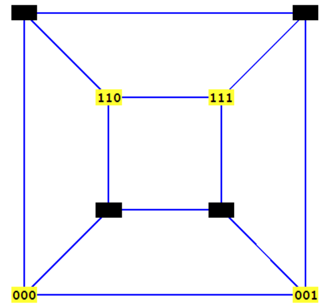
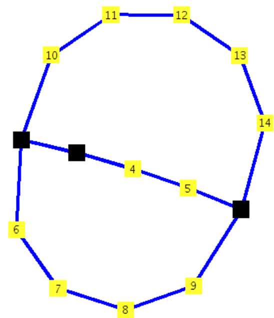
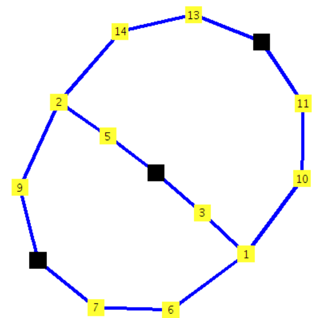
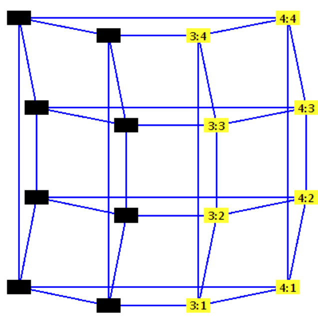
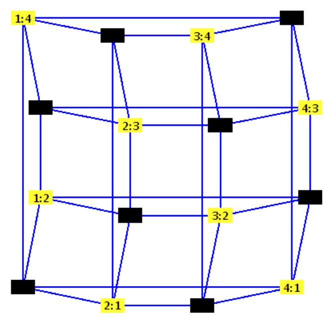
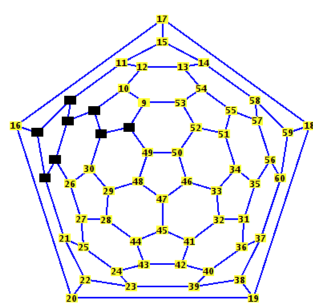
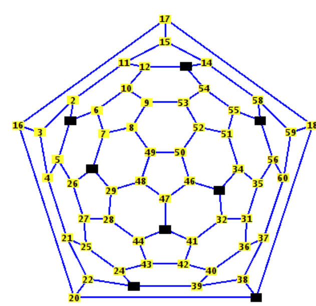
2. The set $\{x \in S_{wd} : x \in \mathbb{R}\}$ is dense in the negative real line.

In conclusion, we have motivated a new family of sets of vertices in graphs which seem to represent sets that are "spread apart". The sets are a generalization of maximally even sets, and follow a physical energy law as if the vertices in the sets were charged particles. We showed that finding such sets, or more precisely finding the minimum k-set energy, is an NP-complete problem. However, we were still able to find bounds and approximations for well-distributed sets and the minimum k-set energy. In special families of graphs we were able to find the well-distributed sets. Similarly, we were able to classify well-distributed sets of size 1,2,3 and $n-1$ (for a graph of order n). We finally showed numerous approaches for further research. Given the physical interpretation of well-distributed sets, it seems practical to analyse and understand them.

Appendix A

Examples of Spread Algorithm

Graph	Initial configuration	Final configuration	Well-distributed?
$P_6 \times P_6$			Unsure
C_{12}			Yes
P_{10}			Yes
Complete binary tree			No

Graph	Initial configuration	Final configuration	Well-distributed?
3D Hypercube			No
$\Theta(4,5,6)$			Yes
$C_4 \times C_4$			Unsure
Soccer Ball Graph			Unsure

Bibliography

[1] Robert B. Allan and Renu Laskar. *On domination and independent domination numbers of a graph.* Discrete Mathematics, Volume 23, issue 2, pp 73-76, 1978.

[2] Béla Bollabás. *Random Graphs.* Academic Press, 1985.

[3] John Clough and Jack Douthett. *Maximally Even Sets.* Journal of Music Theory, Vol. 35, No. 1/2, pp 93-173, 1991.

[4] Karl Dilcher and Sinai Robins. *Zeros and Irreducibility of Polynomials with GCD Powers as Coefficients.* Springer Science+Business Media, 2014.

[5] F M Dong, K M Koh, and K L Teo. *Chromatic Polynomials and Chromaticity of Graphs.* World Scientific, 2004.

[6] Jack Douthett. *The Theory of Maximally and Minimally Even Sets, the One-Dimensional Antiferromagnetic Ising Model, and the Continued Fraction Compromise of Musical Scales.* PhD Thesis, University of New Mexico, 1999.

[7] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, 1979.

[8] Jonathan L. Gross and Jay Yellen. *Graph Theory and its Applications.* Chapman & Hall/CRC, 2006.

[9] Sudipto Guha and Samir Khuller. *Greedy Strikes Back: Improved Facility Location Algorithms.* Journal of Algorithms, Vol. 31, issue 1, pp 228-248, 1999.

[10] Krzysztof Krzywdziński. *Efficient Construction of (d+1,3d)-Ruling Set in Wireless Ad Hoc Networks.* N.T. Nguyen et al. (Eds.): New Challenges in Compu. Collective Intelligence, SCI 244, pp. 147156, 2009.

[11] J.J. Thompson. *XXIV. On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure.* The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, sixth series, pp 237-265, 1904.