Random spanning trees on homogeneous graphs

by

Anna Melikyan

B.S., Yerevan State University, 2012

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Abstract

Spanning trees have been studied as combinatorial objects for a long time. However, they have found great use in other disciplines recently. Disjoint spanning trees are important in fault-tolerant broadcasting or load-balancing communication systems in interconnection networks. For instance, providing completely independent spanning trees (pairwise edge-disjoint and internally disjoint spanning trees) guarantees uninterrupted service. Completely independent spanning trees were introduced by T. Hasunuma and then have been studied on different classes of graphs. The problem to determine whether there exist two completely independent spanning trees in a graph G is NP-hard. In this context, one may aim to construct spanning trees that collide as little as possible. This problem has been studied by Albin, Poggi-Corradini, et al.. Here we think of spanning trees of a graph as random variables and we aim to find a probability mass function (pmf) that minimizes the expected overlap of two random spanning trees.

We study the minimum expected overlap problem for a special class of graphs. We provide a necessary and sufficient condition for a homogeneous graph to admit optimal weights. Given an unweighted graph G, this entails obtaining a weighted graph on the same vertex-set so that the weighted uniform spanning tree pmf is optimal for the minimum expected overlap problem. We use the maximum entropy problem to show this result and apply gradient descent method for the dual objective function to find the optimal weights.

In addition, we show that the objective function of the minimum expected overlap problem is quasiconvex in each coordinate. Essentially, we show that there is a unique minimizer which allows us to introduce convergent algorithms to compute the optimal weights.

Random spanning trees on homogeneous graphs

by

Anna Melikyan

B.S., Yerevan State University, 2012

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Pietro Poggi-Corradini

Copyright

Abstract

Spanning trees have been studied as combinatorial objects for a long time. However, they have found great use in other disciplines recently. Disjoint spanning trees are important in fault-tolerant broadcasting or load-balancing communication systems in interconnection networks. For instance, providing completely independent spanning trees (pairwise edge-disjoint and internally disjoint spanning trees) guarantees uninterrupted service. Completely independent spanning trees were introduced by T. Hasunuma and then have been studied on different classes of graphs. The problem to determine whether there exist two completely independent spanning trees in a graph G is NP-hard. In this context, one may aim to construct spanning trees that collide as little as possible. This problem has been studied by Albin, Poggi-Corradini, et al.. Here we think of spanning trees of a graph as random variables and we aim to find a probability mass function (pmf) that minimizes the expected overlap of two random spanning trees.

We study the minimum expected overlap problem for a special class of graphs. We provide a necessary and sufficient condition for a homogeneous graph to admit optimal weights. Given an unweighted graph G, this entails obtaining a weighted graph on the same vertex-set so that the weighted uniform spanning tree pmf is optimal for the minimum expected overlap problem. We use the maximum entropy problem to show this result and apply gradient descent method for the dual objective function to find the optimal weights.

In addition, we show that the objective function of the minimum expected overlap problem is quasiconvex in each coordinate. Essentially, we show that there is a unique minimizer which allows us to introduce convergent algorithms to compute the optimal weights.

Table of Contents

List of Figures

Acknowledgments

I thank my advisor Professor Pietro Poggi-Corradini for patience, kindness and mathematical ideas. I thank my committee member Dr. Nathan Albin for helping me to see the connection between the maximum entropy problem and weighted uniform spanning trees. I give my thanks to my committee member Dr. Mike Higgins for helpful comments on my dissertation.

Chapter 1

Introduction

Given a graph G, we consider a minimization problem over the set of probability distributions defined on the family of spanning trees of G, where the objective function is the expected pairwise overlap of identically distributed independent spanning trees. As we know the optimal solution for the minimum expected overlap problem exists and can be computed by using the hierarchical structure of non-homogeneous graphs. This entails decomposing any non-homogeneous graph into its homogeneous components. Then, to compute the optimal pmf for the original graph, one may couple the optimal pmfs for the homogeneous components. Moreover, the spanning tree modulus algorithm can be applied to find the optimal pmfs. In this dissertation we discuss the minimum expected overlap problem for homogeneous graphs. This, in a way, is the next step to constructing an optimal pmf for any non-homogeneous graph, as, essentially, homogeneous graphs are the building blocks for constructing an optimal pmf for any graph. We will provide the precise definition of homogeneity of graphs in this context, in the section below.

It is known that any finite multigraph with no self loops has a vertex-induced proper subgraph with the property that every spanning tree of the original graph that is in the support of an optimal pmf restricts to a spanning tree on that subgraph [1]. This special subgraph is called a homogeneous core of a graph. Homogeneous cores are the densest subgraphs of the graph. Graphs that are at least as dense as all their subgraphs are called

homogeneous. Given a non-homogeneous graph, one can identify its densest subgraph fairly cheaply, and then perform a process called deflation, which entails shrinking the homogeneous core into a single vertex. The resulting graph is called a quotient graph. One may apply this technique repeatedly to extract the densest subgraphs from each quotient graph arising at each step of deflation. Iterating this process yields a deflation sequence which allows us to decompose any non-homogeneous graph into its homogeneous components. It is known that to construct an optimal pmf for a non-homogeneous graph one may couple the optimal pmfs for the homogeneous components. Therefore, understanding the construction of the optimal pmfs on homogeneous graphs is essential. As the densest subgraph problem can be solved in polynomial time, the core decomposition allows us to obtain a more efficient way of constructing optimal pmfs for the minimum expected overlap problem.

We categorize homogeneous graphs according to denseness. If a homogeneous graph contains a subgraph with the same denseness, then we call such a graph reducible, otherwise it is called irreducible. This setting allows us to expand the definition of the deflation process to homogeneous reducible graphs. As a result, one may split a homogeneous reducible graph into its homogeneous irreducible components. Similarly, in this case the optimal measure on a homogeneous graph can be constructed by coupling the optimal measures on its irreducible components. Thus, to find an optimal measure for a homogeneous graph, one may couple the optimal measures on irreducible components. This statement leads to the question - how do we construct an optimal pmf for a homogeneous irreducible graph?

We classify spanning trees based on the criterion if they are supported by an optimal pmf or not. The trees that are in the support of an optimal measure are called fair trees. Naturally, we are interested in generating all such trees. There are various fast algorithms that generate uniform spanning trees. We aim to discuss conditions on graphs which would allow us to generate fair trees by using these algorithms.

Precisely our contributions are as follows:

Contributions

- Given a homogeneous graph, we provide a necessary and sufficient condition for the

existence of a weighted uniform spanning tree (WUST) that is optimal for MEO problem.

  - Our proof relies on solving the maximum entropy problem over the set of optimal pmfs for the MEO problem.
  - This allows us to represent our problem as a convex optimization problem.
  - We can compute the optimal weights by applying gradient descent method for the dual problem of the maximum entropy problem.

- We show that the objective function of MEO problem is quasiconvex in each coordinate for homogeneous irreducible graphs.

  - We use Kirchhoff's classical result about the weighted uniform spanning trees and the edge-usage probabilities to translate the MEO problem into the language of effective resistances.
  - This transition allows us to represent the objective function via the pseudo-inverse of the Laplacian of the graph. Finally, we provide algorithms that generate the optimal weights.

1.1 Minimum expected overlap problem

Let $G = (V, E)$ be a finite, connected multigraph with no self-loops, with vertex set V and edge set E. Let μ be a probability mass function (pmf) defined on the family of spanning trees of G, which we denote by Γ_G. The optimal choice of μ allows us to construct spanning trees that on average collide as little as possible. Let $T \in \Gamma_G$ be a spanning tree of G and let $\mathcal{N}$ be a $|\Gamma_G| \times |E|$ matrix with entries $\mathcal{N}(T, e) = \mathbb{1}_{\{e \in T\}}$. In this context, we introduce the *edge-usage probability* which measures how likely it is that a μ-random tree uses a particular edge $e \in E$:

$$\eta(e) = \mathbb{P}_\mu(e \in \underline{T}) = \sum_{T \in \Gamma_G} \mathcal{N}(T, e)\mu(T).$$

The expected overlap of two iid random spanning trees can be computed as follows (see 3.1.1):

$$\mathbb{E}_\mu(\underline{T} \cap \underline{T'}) = \sum_{e \in E} \mathbb{P}_\mu(e \in \underline{T})^2.$$

As we will see in Section 3.1, the minimization problem of the expected overlap of two random independent identically distributed spanning trees can be reduced to the problem of minimizing the variance of the edge-usage probabilities for a given graph $G = (V, E)$. Since the number of edges is always the same for all spanning trees of G, and is equal to $|V| - 1$, one may use this fact to show that solving the MEO problem is equivalent to solving the following fairest edge-usage (FEU) problem

$$\begin{aligned} &\text{minimize} && \mathrm{Var}(\eta) \\ &\text{subject to} && \mathbb{P}_\mu(e \in \underline{T}) = \eta(e) \quad e \in E. \end{aligned}$$

Therefore, if there is no discrepancy in utilization of edges between any pairs of μ-random trees, then μ minimizes their collision on average as well. That being said, we introduce a class of graphs for which it is possible to select trees so that they all use the edges of the graph fairly. We call such graphs *homogeneous graphs*. It is known that every graph can be partitioned into homogeneous components. Moreover, the optimal measure can be decomposed into the optimal measures on homogeneous components. In a way, homogeneous graphs are building blocks for constructing an optimal measure. If G is homogeneous, then the edge-usage probabilities are the reciprocal of the *denseness* of the graph which is the following quantity:

$$\theta(G) = \frac{|E|}{|V| - 1}.$$

Essentially, it tells us how dense or sparse the graph is. If a homogeneous graph has a vertex-induced, connected subgraph with the same denseness (homogeneous graphs with

such property are called *reducible*, if there is no such subgraph, then we call the graph *irreducible*), then we partition it until we are left with irreducible components. The optimal measure on a homogeneous graph can be decomposed into the optimal measures on its irreducible components. Thus, to find an optimal measure for a homogeneous graph, one may couple the optimal measures on irreducible components (see Theorem 3.4.2). We desire to generate spanning trees that are in the support of an optimal measure (we call such trees *fair trees* since they use the edges of the graph fairly).

1.2 Weighted uniform spanning trees

The theory of random walks lies behind many algorithms that generate uniform spanning trees (UST). A weighted uniform spanning tree is a random spanning tree arising, as a union of first-visited edges, from a weighted random walk on a weighted graph $G = (V, E, \sigma)$, where $\sigma \in \mathbb{R}^E_{\geq 0}$, or equivalently, a random tree T with probability $\mu_\sigma(T)$ proportional to its weight:

$$\prod_{e \in T} \sigma(e).$$

Wilson's algorithm (1996) [40] gives rise to a random spanning tree that has distribution μ_σ. It uses the notion of loop erasure of a path which had been studied by Lawler in the 1980's. We would like to use Wilson's algorithm to generate fair trees for a weighted homogeneous irreducible graph. Having in mind this option, we consider the following MEO problem for a weighted graph:

$$\begin{aligned} \text{minimize} \quad & \sum_{e \in E} \mathbb{P}_{\mu_\sigma}(e \in T)^2 \\ \text{subject to} \quad & \sigma \in \mathbb{R}^E_{\geq 0}, \quad \mu_\sigma \in \mathcal{P}(\Gamma) \end{aligned}$$

In Chapter 4 (see Theorem 4.4.8), we will show the following.

Theorem 1.2.1. *A homogeneous graph G is irreducible if and only if there exists σ such that the WUST μ_σ is optimal for the* MEO *problem.*

In Chapter 2, we recall Kirchhoff's result which establishes a connection between the edge-usage probabilities and the effective resistances on edges for electrical networks with edge-conductances given by σ, i.e., if $e = \{x, y\} \in E$, then

$$\mathbb{P}_{\mu_\sigma}(e \in T) = \sigma(x, y)\mathcal{R}_{\text{eff}_\sigma}(x, y).$$

Therefore, we are concerned with the following objective function in σ.

$$\mathcal{E}(\mu_\sigma) = \sum_{e \in E} \mathbb{P}_{\mu_\sigma}(e \in T)^2 = \sum_{e \in E} \sigma(x, y)^2 \mathcal{R}_{\text{eff}_\sigma}(x, y)^2,$$

where we refer to $\mathcal{E}(\mu_\sigma)$ as the energy of μ_σ.

We compute the effective resistance between s and t by the following formula :

$$\mathcal{R}_{\text{eff}}(s, t) = (\delta_t - \delta_s)^T L^+ (\delta_t - \delta_s),$$

where L^+ is the pseudo inverse of the Laplacian of the graph and δ_i is the i-th unit vector. To solve the optimization problem we apply rank-one update for the Laplacian.

1.3 Algorithms

In Chapter 5 we discuss algorithms that generate optimal weights for the weighted MEO problem. Both of the algorithms update the weights on edges based on theoretical analysis done for the optimal α (see section 5.5). The first algorithm is based on a coordinate descent method. We start out with initial weights, then update the weight on each edge optimally one at a time until the change in energy is sufficiently small.

The second algorithm is based on the argument of equalizing per-edge effective resistances. This entails selecting the edge with the smallest per edge effective resistance and updating it optimally. Finally, we record the change in energy and stop the process when it is sufficiently small.

Chapter 2

Uniform spanning trees

A tree is a connected graph without cycles. Given a connected graph, a spanning tree is a subgraph that is also a tree and contains every vertex of the given graph. We will be discussing random spanning trees of finite connected graphs in this chapter. A given graph may have a large number of spanning trees. How do we select one uniformly at random ? In section 2.2 we will discuss algorithms that generate uniform spanning trees. But first in section 2.1 we give a more formal definition of a uniform spanning tree as a special case of a weighted uniform spanning tree which, too, will be discussed in section 2.1.

2.1 Weighted uniform spanning trees (WUST)

Let $G = (V, E, \omega)$ be a finite connected graph with weights ω. We consider the family of spanning trees on G. We think of spanning trees as random variables with distribution μ.

Definition 2.1.1. We say that T is a weighted uniform spanning tree (WUST) if it is a random tree with distribution proportional to its weight, i.e., $\prod_{e \in T} \omega(e)$. More formally,

$$\mu_\omega(T) = \frac{\prod_{e \in T} \omega(e)}{\sum_{T' \in \Gamma_G} \prod_{e \in T'} \omega(e)}.$$

Remark 2.1.2. In case when $\omega(e) = 1$ for all $e \in E$, T is a uniform spanning tree.

In other words, we are considering the set of spanning trees of a given graph G. Then we choose one of them at random with probability equal to the reciprocal of the size of that set. This random choice of T makes it a *uniform random spanning tree*.

2.2 Generating uniform spanning trees (UST)

We can classify the algorithms generating Uniform Spanning Trees into two main categories - random walk-based algorithms and determinant-based algorithms. The earliest algorithms are based on Kirchhoff's well-known Matrix-Tree theorem, which allows us to compute the number of spanning trees of a given graph by computing the determinant of a submatrix of the Laplacian of that graph. The running time of the first such algorithm is $O(mn^3)$. Even though this result had been improved over the years, random walk-based algorithms are known to generate uniform spanning trees faster. Therefore, we will discuss them in more detail in this section. More details about the algorithms can be found in [24].

2.2.1 Aldous-Broder algorithm

The expected running time for Aldous-Broder (1989) algorithm is $O(nlogn)$ for each generated tree for almost all graphs, and $O(n^3)$ for the worst graphs. The simulation of the simple random walk on a connected undirected graph allows us to generate a spanning tree uniformly at random. Here is how the algorithm works.

Aldous-Broder Algorithm

1. Simulate a simple random walk on a graph $G = (V, E)$ starting at an arbitrary vertex v_0 until we visit all vertices. For every vertex $v_i \in V \setminus v_0$ keep record of the edge (v_j, v_i) which allows us to visit v_i for the first time.

2. Output the resulting collection of edges.

Aldous and Broder showed independently that this method gives rise to a uniform spanning tree. The expected running time of this algorithm is the cover time of the simple

random walk. It is known that for any connected graph the expected cover time is $O(n^3)$ [4]. However it was shown that if the second eigenvalue (algebraic connectivity of the graph) of the adjacency matrix (transition probability matrix for the simple random walk) is bounded above by 1, then the cover time is $O(n \log n)$ [10]. This condition is true for almost every graph (Erdös-Rényi $G_{n,p}$ with $p > c \log n/n$) [19], and for almost all $d-$regular graphs [9], [18].

2.2.2 Wilson's algorithm

Wilson's algorithm (1996) is the fastest method known for generating uniform spanning trees. The expected running time of this algorithm is proportional to the mean hitting time of the simple random walk. At the heart of this algorithm lies the idea of a *loop erasure* of a path coined and developed by Lawler in 1980.

Let $\mathcal{P}$ be a finite path $< x_0, x_1, ..., x_l >$ in G. We obtain a *loop erasure* of $\mathcal{P}$: $\mathrm{LE}(\mathcal{P}) =< u_0, u_1, ..., u_m >$ by erasing cycles in $\mathcal{P}$ in the order they appear. More formally, let $u_0 = x_0$. If $x_l = u_0$, we set $m = 0$ and terminate the process. Otherwise, we set u_1 be the fist vertex in $\mathcal{P}$ after we visit x_0, i.e., $u_1 = x_{i+1}$, where $i = \max\{j : x_j = x_0\}$. If $x_l = u_1$, we set $m = 1$ and we terminate the process. Otherwise, we let u_2 be the first vertex in $\mathcal{P}$ after the last visit to u_1, and so on. Wilson's algorithm creates a sequence of trees by loop-erasing the paths until we uncover a spanning tree. Wilson has shown that this method gives rise to a uniform spanning tree [40]. Let r be any vertex from V. We enumerate the vertices from $V \setminus \{r\}$. Let $T(0) := \{r\}$, then we obtain $T(1)$ by picking a vertex v_j at random from $V \setminus \{r\}$ and run a loop-erased random walk until we hit $T(0)$, i.e., we append the loop-erasure of the path arising from a simple random walk starting at v_j till we hit $T(0)$, to $T(0)$. Now assume $T(i)$ is known. Then $T(i+1) = T(i) \cup \mathrm{LE}(\text{path from } v_{i+1} \text{ to } T(i))$. More formally, we will have

1. $T(0) := \{r\}, < v_1, ..., v_{n-1} >= V \setminus \{r\}$.

2. Suppose $T(i)$ has been generated.

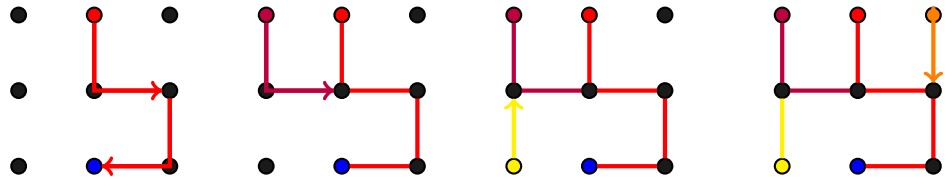


Figure 2.1: *Wilson's Algorithm applied to a 3 by 3 grid*

3. Start a random walk at v_{i+1} and stop when it hits $T(i)$.

4. $T(i+1) = T(i) \cup \text{LE}(\text{path from } v_{i+1} \text{ to } T(i))$

2.2.3 Wilson's algorithm for grids

In this section we illustrate Wilson's algorithm on an example for a 3 by 3 grid as in Figure 2.1. First, we enumerate the vertices of the grid. We start a random walk from a randomly selected vertex colored red until we hit the vertex colored blue and we loop erase it to form $T(1)$. Essentially, $T(1) = \text{LE path (red} \rightarrow \text{blue)}$. Similarly, we run a random walk from the next selected vertex until we hit $T(1)$, then we loop erase the walk and append the resulting path to $T(1)$. Thus forming a new loop-erased path $T(2)$. We continue this process till we uncover a tree. As we have discussed in the previous section this method gives rise to a uniform spanning tree.

2.3 Electrical networks and WUST

In this section we will briefly discuss electrical circuit theory. Given a resistor network we can present it as a graph with current sources as vertices and conductances as edge-weights. In section 2.3.1 we define flows and effective resistance on graphs. Then we discuss the connection between the effective resistance of the graph and the Laplacian in section 2.3.3. Finally, we discuss Kirchhoff's theorem which establishes a connection between uniform spanning trees and effective resistances of graphs.

2.3.1 Electrical current flows and effective resistance

Let $G = (V, E)$ be a finite simple graph with vertex-set V and edge-set E. Choose an orientation on each edge, and call the oriented edge-set $\vec{E}$. Also let $\hat{E} = \vec{E} \cup \vec{E}_R$, where $\vec{E}_R$ is the set of edges with the opposite orientation of $\vec{E}$.

Definition 2.3.1. Let s and t be two different nodes on G so that s is the source and t is the sink. A *flow* from s to t is a function $f : \hat{E} \to R$ such that

(i) f is antisymmetric : $f(x, y) = -f(y, x)$.

(ii) f satisfies the node law: For every $x \in V \setminus \{s, t\}$

$$\mathrm{Div}_f(x) := \sum_{y \sim x} f(x, y) = 0.$$

Definition 2.3.2. The *value* of a flow from s to t is

$$\mathrm{Val}(f) := \mathrm{Div}_f(s)$$

Given a flow from s to t and some edge-resistances $r(e) > 0$ for $e \in E$, let the energy of the flow be

$$\mathcal{E}(f) = \sum_{e \in E} r(e)|f(e)|^2.$$

We denote the minimum energy among all flows from s to t of value 1 by $\mathcal{E}^*$. It is known that if f is a unit current flow from s to t with potential h ,then

$$\mathcal{E}^* = \sum_{e \in E} r(e)|f(e)|^2 = h(t) - h(s).$$

If we replace the whole network by a single edge $\hat{e}$ with resistance $\mathcal{R}$ between s and t , but we keep the unit flow and the potential difference as before ,then $\mathcal{R}$ would have to satisfy :

$$\mathcal{R}(\hat{e})|f(\hat{e})|^2 = \mathcal{R}(\hat{e}) = h(t) - h(s).$$

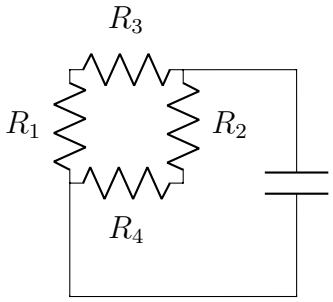


Figure 2.2: *Square: Electrical Circuit Interpretation*

We call $\mathcal{R}$ the effective resistance of the network from s to t and write $\mathcal{R}_{\text{eff}}(s,t)$.

Here we discuss an example of a square as an electrical circuit as in Figure 2.2. We assume $R_1 = R_2 = R_3 = R_4 = 1$. Then the parallel and serial rules imply

$$\mathcal{R}_{\text{eff}}(e_1) = \frac{R_1R_2 + R_1R_3 + R_1R_4}{R_1 + R_2 + R_3 + R_4} = \frac{3}{4}.$$

2.3.2 Potential function and electrical circuits

Alternatively, one may approach the electrical circuit theory by starting out from defining a potential function. In this section we will adopt this method.

Let $G = (V, E)$ be a finite graph with no loops or multiple edges. Let $c : E \to (0, \infty)$ be a weight function on the edges. G is an electrical network with the edge *conductances* given by function c.

Definition 2.3.3. Given a source s and a sink t in an electrical network, the *voltage* h between s and t is a harmonic function on $V \setminus \{s, t\}$, i.e., for any $x \in V \setminus \{s, t\}$

$$h(x) = \sum_{y \in V \setminus \{x\}} h(y)p(x, y),$$

Moreover, we define $h(s) = 0$ and $h(t) = 1$.

To continue evolving the story we want to share, we would like to present the connection between finite Markov chains and electrical networks, first.

Definition 2.3.4. A sequence of random variables $(X_0, X_1, ...)$ is a *Markov chain* with state space Ω and transition matrix P if for all $x, y \in \Omega$, all $t \geq 1$, and all events $H_{t-1} = \cap_{s=0}^{t-1}\{X_s = x_s\}$ satisfying $P(H_{t-1} \cap \{X_t = x\}) > 0$, we have the following *Markov property*

$$P\{X_{t+1} = y \mid H_{t-1} \cap \{X_t = x\}\} = P\{X_{t+1} = y \mid X_t = x\} = P(x, y).$$

Electrical networks provide an alternative language for reversible Markov chains. We will consider a weighted random walk on the nodes of the network.

Definition 2.3.5. Given a graph $G = (V, c(E))$, *simple random walk* on G is the Markov chain with state space V and transition matrix

$$P(x, y) = \begin{cases} \frac{c(x,y)}{\sum_{z \sim x} c(x,z)} & \text{if } y \sim x \\ 0 & \text{otherwise} \end{cases}$$

In addition, in this section, we will be referring to irreducible Markov chains.

Definition 2.3.6. A chain P is called *irreducible* if for any two states $x, y \in \Omega$ there exists integer t such that

$$P^t(x, y) > 0.$$

To show the existence and uniqueness of the potential function we will refer to the following theorem.

Theorem 2.3.7. *Let (X_t) be a Markov chain with irreducible transition matrix P, let $B \in \Omega$, and let $h_B : B \to \mathbb{R}$ be a function defined on B. The function $h : \Omega \to \mathbb{R}$ defined by $h(x) = \mathbb{E}_x(h_B(X_{\tau_B}))$ is the unique extension $h : \Omega \to \mathbb{R}$ of h_B such that $h(x) = h_B(x)$ for all $x \in B$ and h is harmonic for P at all $x \in \Omega \setminus B$.*

Below we provide the formal definition of a hitting time .

Definition 2.3.8. Given a Markov chain with state space Ω, the *hitting time* τ_B of a subset $B \in \Omega$ is the first time the chain visited one of the nodes in B. If (X_t) is a random walk, the

hitting time of B is given by

$$\tau_B := \min\{t \geq 0 : X_t \in B\}.$$

2.3.3 Laplacian of a graph

In the previous section we discussed the transition matrix of a weighted random walk. In this section we discuss a different matrix associated with any graph $G = (V, c(E))$. Here, we provide definitions for the Laplacian matrix and the pseudo-inverse of the Laplacian.

Let $c : E \to \mathbb{R}_{>0}$ be edge-conductances for G. The Laplacian $L : V \times V \to \mathbb{R}$ is defined as follows:

$$L(x, y) = \begin{cases} C(x) = \sum_{z \sim x} c(x, z) & x = y \\ -c(x, y) & x \neq y \end{cases}$$

Let $h : V \to \mathbb{R}$ be a vector in $\mathbb{R}^N$. Then we can apply Laplacian operator to h which yields the following formula for $Lh(x)$ for $x \in V$:

$$(Lh)(x) = \sum_{y \in V} L(x, y)h(y) = C(x)h(x) - \sum_{y \sim x} c(x, y)h(y).$$

The spectrum of the Laplacian matrix of a graph plays a vital role in analysis of its topological characteristics such as minimal cuts, clustering and the number of spanning trees [6]. Moreover, the pseudo-inverse of the Laplacian has found great use in many fields such as probability and mathematical chemistry, collaborative recommendation systems and social networks, epidemiology and infrastructure planning [17, 25, 27, 29, 33, 35, 42].

Despite its theoretical high importance, in practice the computation of the pseudo-inverse of a Laplacian is costly. Typically, it takes $O(n^3)$ computational time to compute it by applying standard matrix factorization and inversion based methods. For instance, for dynamic time-evolving graphs with many vertices, such as online social networks, this computational complexity is undesirable. However, there has been an improvement for this time if one inherits a divide-and-conquer algorithm based approach introduced in [36]. The divide oper-

ation in this case requires determining an arbitrary bi-partition of a graph - a cut of a graph that consists of exactly two connected components resulting from a deletion of k edges. The computational complexity is $O(k \cdot n^2)$ in this case.

It is known that the Laplacian is a symmetric and positive semi-definite matrix. Particularly, it has real and non-negative eigenvalues, hence they can be ordered as $0 \leq \lambda_1 \leq ... \leq \lambda_N$. Let u_j be the orthonormal basis for L. Then we can write L as follows:

$$L = \sum_{j=2}^{N} \lambda_j u_j {u_j}^T.$$

The pseudo-inverse of L can be defined as follows when $\lambda_2 > 0$:

$$L^+ = \sum_{j=2}^{N} \frac{1}{\lambda_j} u_j {u_j}^T.$$

Theoretically, one may compute L^+ through a straight-forward approach by an inversion of its non-zero eigenvalues. In practice, one may use singular value decomposition (SVD) based method, which is fairly expansive as it generates the pseudo-inverse in $O(n^3)$ computational time. Alternatively, one may use a rank-one perturbation of L for a simple, connected, undirected graph, which makes it invertible, hence allows us to compute L^+ as in [36].

2.3.4 Pseudo-inverse of the Laplacian and the effective resistance

There is an interesting analogy between graphs and resistive electrical circuits [17, 29, 16]. For instance, the effective resistance distance has many rich applications [17, 29]. In addition, it determines the expected length of random commutes between node pairs in the graph [11, 39]. Particularly, below we provide the formula connecting the effective resistance of a graph with the pseudo-inverse of a Laplacian.

Let $L^+ = \{l_{ij}^+\}_{i,j=1}^n$, then the effective resistance of the graph G can be computed as follows:

$$\mathcal{R}_{\text{eff}}(x, y) = l_{xx}^+ + l_{yy}^+ - 2l_{xy}^+. \tag{2.3.1}$$

In other words, we may often write this statement as follows. Let δ_i be the i-th unit vector. Let $u = \delta_y - \delta_x$ for some $x, y \in V$. Then the effective resistance from x to y can be represented as :

$$\mathcal{R}_{\text{eff}} = u^T L^+ u.$$

2.3.5 Kirchhoff's theorem

Finally, we discuss Kirchhoff's theorem [26] (1847) in this section.

Theorem 2.3.9. *Let $G = (V, E, \omega)$ be a graph where we think of the edge-weights ω as edge-conductances c. Then for a given edge $e \in E$ the weighted uniform spanning trees are connected with effective resistances by the following formula:*

$$\mathbb{P}_{\mu_\omega}(e \in T) = c(e)\, \mathcal{R}_{\mathit{eff}}(e),$$

where the left hand side represents the edge-usage probability which measures the likelihood of e appearing in a weighted uniform spanning tree. Particularly, we can define it as follows

$$\mathbb{P}_{\mu_\omega}(e \in T) = \sum_{T \in \Gamma_G} \mu_\omega(T) \mathcal{N}(T, e),$$

where $\mathcal{N}(T, e) = \mathbb{1}_T(e)$ is the indicator function.

More detailed explanation for this result can be found in [31] and [15]. Particularly, if we compute the edge-usage probabilities in the case of the square as in Figure 2.2, we would have the following for every $e \in E$

$$\mathbb{P}_{\mu_\sigma}(e \in \underline{T}) = \frac{\#\text{ of spanning trees that contain } e}{\#\text{ of spanning trees}} = \frac{3}{4}.$$

Theorem 2.3.10. *Let T be a spanning tree chosen uniformly at random from the set of spanning trees of G. Then the unique solution of the Kirchhoff's laws with source s and sink*

t, and size 1 *is given by*

$$i(x,y) = \mathbb{P}(T \textit{ has the property that the unique path } s/t \textit{ passes along the edge } \{x,y\})$$
$$- \mathbb{P}(T \textit{ has the property that the unique path } s/t \textit{ passes along the edge } \{y,x\}).$$

2.4 Counting spanning trees

In this section we will be discussing various methods to compute the number of spanning trees of a given graph. In section 2.4.1 we discuss Kirchhoff's well known Matrix-Tree theorem. In section 2.4.2 we discuss deletion and contraction of graphs where these operations allow us to reduce the graph into smaller graphs. Often this leads to obtaining a recursion formula for the number of spanning trees for the original graph.

2.4.1 Matrix-Tree theorem

In this section we discuss the well-known Matrix-tree theorem. One may find the proofs of these results in the lecture notes by N. Srivastava [38].

Let $A[i]$ be the matrix A with its i-th column and row removed.

Theorem 2.4.1. *For a given graph G with Laplacian L_G, the number of spanning trees can be computed by finding the determinant of a submatrix of the Laplacian.*

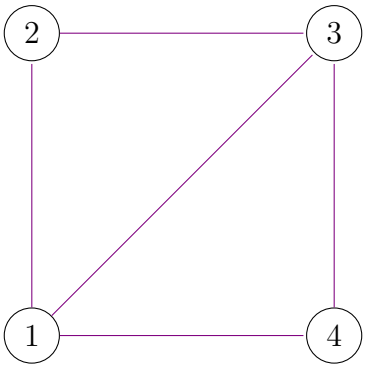
$$\tau(G) = \det(L_G[i]).$$

However, often in literature matrix-tree theorem is associated with spectral representation of the determinant of the Laplacian via its eigenvalues. Hence, we will discuss this formulation here as well.

Theorem 2.4.2 ([7])**.** *Let $0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ be the eigenvalues of L_G. Then*

$$\tau(G) = \frac{1}{n} \prod_{i=2}^{\infty} \lambda_i.$$

Here we consider the example of square with diagonal.



The associated Laplacian with this graph would be:

$$L_G = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$

If we remove the fourth row and column, we will get the following matrix:

$$L_G[4] = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix}$$

Calculations yield that $\det(L_G[4]) = 8$. By Matrix-Tree theorem we conclude that $\tau(G) = 8$, which could be easily verified by counting.

2.4.2 Deletion-Contraction theorem

We will introduce a few concepts related to counting the number of spanning trees $\tau(G)$ for any graph G. Deletion and contraction are basic graph operations defined as follows.

Definition 2.4.3. Let G be a graph and $e \in E(G)$ an edge. The *deletion* $G - e$ is the graph obtained by erasing e, leaving its endpoints and everything else intact. The *contraction* G/e is obtained by erasing e and merging its endpoints into a single vertex.

It turns out that deletion and contraction can be used to compute the number of spanning trees for a graph by removing one edge at a time.

Theorem 2.4.4. *If $e \in E(G)$ is not a self-loop, then $\tau(G) = \tau(G - e) + \tau(G/e)$.*

We desire to compute the number of spanning trees for the N-story House. An N-story House is formed from an $(N-1)$-story House by adding a *pavement* floor and two vertical walls at the bottom. In general a *floor* will be comprised of an horizontal edge and the two walls standing on top of it on each side. Finally, the triangle at the top is called the *roof*. We denote the resulting graph as H_N.

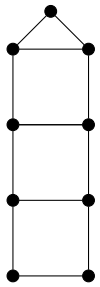


Figure 2.3: 3-*storey House*

We apply Theorem 2.4.4 to the N-story House to compute $\tau(H_N)$. First we delete and contract the pavement edge p on the bottom floor. Note that

$$\tau(H_N - p) = \tau(H_{N-1}),$$

because the two walls that where connected to the pavement will have to belong to any spanning tree of $H_N - p$.

On the other hand $(H_N)/p$ is a double-roofed graph with $N-1$ floors which we denote by $\hat{H}_N$. It will be useful to keep track of both $\tau_N := \tau(H_N)$ and $\delta_N := \tau(\hat{H}_N)$. We have just shown that

$$\tau_N = \tau_{N-1} + \delta_N. \tag{2.4.1}$$

We now study the double-roofed House $\hat{H}_N$ and δ_N. Again, we apply Theorem 2.4.4 to an edge e on the bottom roof. As before,

$$\tau(\hat{H}_N - e) = \tau(H_{N-1}) = \tau_{N-1}.$$

However, $(\hat{H}_N)/e$ is an $(N-1)$-story House $\tilde{H}_{N-1}$ with a double pavement, meaning that the bottom edge is a multi-edge of multiplicity 2. Therefore, we apply Theorem 2.4.4 one more time to one of the two bottom edges of $\tilde{H}_{N-1}$ that we call f. On one hand,

$$\tau(\tilde{H}_{N-1} - f) = \tau_{N-1}.$$

On the other hand, contracting f gives rise to a double-roofed house $\hat{H}_{N-1}$ with a self-loop at the tip of the bottom roof. This self-loop can be removed because it won't appear in any spanning tree. Summarizing we have found that

$$\delta_N = 2\tau_{N-1} + \delta_{N-1}. \tag{2.4.2}$$

Combining (2.4.3) and (2.4.2) we get the following recurrence:

$$\begin{bmatrix} \tau_N \\ \delta_N \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} \tau_{N-1} \\ \delta_{N-1} \end{bmatrix}$$

Note that H_1 is the regular House graph, so $\tau_1 = 11$, while $\hat{H}_1$ is the square with one diagonal, so $\delta_1 = 8$. Iterating $N-1$ times we obtain the following equation.

$$\begin{bmatrix} \tau_N \\ \delta_N \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}^{N-1} \begin{bmatrix} \tau_1 \\ \delta_1 \end{bmatrix}$$

Set

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}.$$

We will diagonalize A to compute its powers. We want to write

$$A = S\Lambda S^{-1},$$

where Λ is the diagonal matrix with the eigenvalues of A as diagonal entries. Then, we will have

$$A^{N-1} = S\Lambda^{N-1}S^{-1}.$$

A computation shows that

$$S = \begin{bmatrix} \frac{1+\sqrt{3}}{2} & \frac{1-\sqrt{3}}{2} \\ 1 & 1 \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} 1 & \frac{\sqrt{3}-1}{2} \\ -1 & \frac{1+\sqrt{3}}{2} \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 2+\sqrt{3} & 0 \\ 0 & 2-\sqrt{3} \end{bmatrix}$$

Therefore A^{N-1} equals

$$\frac{1}{\sqrt{3}} \begin{bmatrix} \frac{1+\sqrt{3}}{2}(2+\sqrt{3})^{N-1} - \frac{1-\sqrt{3}}{2}(2-\sqrt{3})^{N-1} & (2+\sqrt{3})^{N-1} - (2-\sqrt{3})^{N-1} \\ (2+\sqrt{3})^{N-1} - (2-\sqrt{3})^{N-1} & (1+\sqrt{3})(2+\sqrt{3})^{N-1} - (1-\sqrt{3})(2-\sqrt{3})^{N-1} \end{bmatrix}$$

In particular, τ_N equals

$$11\frac{1+\sqrt{3}}{2\sqrt{3}}(2+\sqrt{3})^{N-1} - 11\frac{1-\sqrt{3}}{2\sqrt{3}}(2-\sqrt{3})^{N-1} + \frac{8}{\sqrt{3}}(2+\sqrt{3})^{N-1} - \frac{8}{\sqrt{3}}(2-\sqrt{3})^{N-1} \quad (2.4.3)$$

and when $N = 2$ we get that $\tau_2 = 41$.

Chapter 3

MEO and FEU problems

In section 3.1 we introduce the MEO and FEU problems. Then we provide a brief summary of results from [1] in sections 3.2 and 3.3. First we define a new class of graphs where every edge has equal probability to appear in a random spanning tree chosen with distribution minimizing the pairwise expected overlap. Then we see that every non-homogeneous graph has a homogeneous core which solves the densest subgraph problem. In addition, we discuss a core decomposition called deflation to find the optimal distributions for each homogeneous core and then couple them to construct an optimal measure for the original MEO problem.

3.1 MEO and FEU problems

Let $G = (V, E)$ be a finite connected graph. Let Γ_G denote the family of spanning trees on G, and T, T' be spanning trees in Γ_G. The overlap of two spanning trees is $|T \cap T'| = \sum_{e \in E} \mathcal{N}(T, e)$, where $\mathcal{N}(T, e)$ is the indicator function of e in T. Let $\mathcal{P}(\Gamma_G)$ be the set of probability mass functions defined on Γ_G. We say that a random spanning tree is μ-random if it has distribution $\mu \in \mathcal{P}(\Gamma_G)$, i.e., $\mathcal{P}_\mu(\underline{T} = T) = \mu(T)$. The expected overlap of two μ-random independent

spanning trees $\underline{T}$ and $\underline{T'}$ can be computed as follows:

$$\begin{aligned}\mathbb{E}_\mu|\underline{T}\cap\underline{T'}| &= \sum_{T,T'\in\Gamma_G}|T\cap T'|\mu(T)\mu(T')\\ &= \sum_{e\in E}\Big(\sum_{T\in\Gamma_G}\mu(T)\mathcal{N}(T,e)\Big)^2\\ &= \mu^T\mathcal{N}\mathcal{N}^T\mu\end{aligned} \tag{3.1.1}$$

.

We desire to find a probability mass function (pmf) that minimizes this overlap. In other words we aim to solve the following Minimum expected overlap (MEO) problem :

$$\begin{aligned}&\text{minimize} \quad \mathbb{E}_\mu|\underline{T}\cap\underline{T'}|\\ &\text{subject to} \quad \mu\in\mathcal{P}(\Gamma_G).\end{aligned}$$

This allows us to rewrite the expected overlap in terms of the edge-usage probabilities.

$$\mathbb{E}_\mu(\underline{T}\cap\underline{T'}) = \sum_{e\in E}\mathbb{P}_\mu(e\in\underline{T})^2. \tag{3.1.2}$$

Let $\eta(e) = \mathbb{P}_\mu(e\in\underline{T})$. Then

$$\begin{aligned}\mathrm{Var}(\eta) = \mathbb{E}(\eta^2) - (\mathbb{E}(\eta))^2 &= \frac{1}{|E|}\sum_{e\in E}\big(\mathbb{P}_\mu(e\in\underline{T})\big)^2 - \left(\frac{1}{|E|}\sum_{e\in E}\sum_{T\in\Gamma_G}\mathcal{N}(T,e)\mu(T)\right)^2\\ &= \frac{1}{|E|}\mathbb{E}_\mu(\underline{T}\cap\underline{T'}) - \left(\frac{1}{|E|}\sum_{T\in\Gamma_G}\mu(T)\sum_{e\in E}\mathcal{N}(T,e)\right)^2\\ &= \frac{1}{|E|}\mathbb{E}_\mu(\underline{T}\cap\underline{T'}) - \left(\frac{|V|-1}{|E|}\right)^2.\end{aligned}$$

To solve the MEO problem for a given graph it is enough to find η for the FEU problem.

$$\begin{aligned} &\text{minimize} \quad \mathrm{Var}(\eta) \\ &\text{subject to} \quad \mathbb{P}_\mu(e \in \underline{T}) = \sum_{\gamma \in \Gamma_G} \mu(\gamma)\mathcal{N}(\gamma, e) = \eta(e) \quad (\forall e \in E). \end{aligned} \tag{3.1.3}$$

The set of solutions for MEO problem is non-empty in $\mathcal{P}(\Gamma_G)$, since we are minimizing a quadratic form over a compact set. How do we construct such optimal pmfs? In the next section we discuss the necessary tools provided in [1] to find such pmfs. As we have established the existence of such optimal pmfs it would be natural to classify the spanning trees based on the fact whether they are in the support of an optimal measure or not.

Definition 3.1.1. Spanning trees are called *fair* if they are in the support of an optimal measure.

$$\Gamma_G^f = \{T \in \Gamma_G \quad | \quad \exists \quad \mu^* \in \mathcal{P}(\Gamma_G) \quad \text{such that} \quad \mu^*(T) > 0\}.$$

If $T \in \Gamma_G \setminus \Gamma_G^f$, then T is called *forbidden.*

3.2 Homogeneous graphs and denseness ratio

As a result of introducing the FEU problem, it is natural to ask a question how easy would it be to find graphs where edge-usage probabilities are the same for all edges of a given graph. This gives rise to studying a new class of graphs where μ-random spanning trees utilize the edges of the graph fairly.

Definition 3.2.1. A graph G is called *homogeneous* if there exists optimal μ for the FEU problem such that $\mathbb{P}_\mu(e \in \underline{T})$ is the same for every edge e. Moreover,

$$\mathbb{P}_\mu(e \in \underline{T}) = \frac{|V| - 1}{|E|}.$$

We call the reciprocal of the ratio above *denseness ratio.* There are various notions of density known in the literature. The denseness ratio we define here and will be working with

in the next sections resembles the average-degree density. More formally, the denseness of a graph $G = (V, E)$ is the following quantity:

$$\theta(G) = \frac{|E|}{|V| - 1}.$$

For instance, spanning trees have denseness ratio equal to 1, while complete graphs have denseness ratio equal to $|V|/2$.

3.2.1 Densest subgraph problem and homogeneous cores

Let $\mathcal{H}$ be the family of vertex-induced proper subgraphs of G. We say that H is the densest subgraph of G, if it solves the following optimization problem:

$$\theta(H) = \max_{K \in \mathcal{H}} \theta(K).$$

It is known that one can detect densest subgraphs for a given graph in polynomial time [22].

There is a linear-programming based approach as well as a linear 2-approximation algorithm that solves the densest subgraph problem [12]. More recently, Bahmani et al. [5] provided a near-optimal algorithm in MapReduce model of computation. In this paper new algorithms for finding the densest subgraph in the streaming model have been presented. For any $\varepsilon > 0$, the algorithms make $O((\log n)/\log(1+\varepsilon))$ passes over the input and find a subgraph whose density is guaranteed to be within a factor $2(1+\varepsilon)$ of the optimum.

Due to the usefulness of dense components, it is generally accepted that their existence is a rule rather than the exception [34]. Below we consider a dense subgraph with a special property defined as follows.

Definition 3.2.2. Let $H \subset E$ be a subgraph of G, we say H has restriction property if every fair tree $T \in \Gamma_G^f$ restricts to a spanning tree of H.

Definition 3.2.3. Let $H \in \mathcal{H}$ be a homogeneous subgraph of G with restriction property. We call H a *homogeneous core* of G.

The homogeneous core of a given graph is the densest subgraph of that graph. Theorem 5.9[1] in conjunction with Corollary 5.10 [1] provide a proof for this. In addition, it is known that every graph contains a homogeneous core (Theorem 5.2[1]).

3.3 Deflation and serial rule

In this section we discuss a core decomposition process called *deflation*, which allows us to decompose a nonhomogeneous graph into homogeneous components. In addition, we will see that we can construct an optimal pmf for such graphs by coupling the optimal pmfs for each homogeneous component. The *deflation process* for a graph G turns its densest subgraph H into a single vertex. The resulting graph G/H is called a *quotient graph* . The vertices of G/H are obtained by identifying all vertices of H in G as a single vertex v_H. More formally,

$$V(G/H) = (V(G) \setminus V(H)) \cup \{v_H\}.$$

The edges in $E \setminus H$ are sent to $E_{G/H}$, while the edges in H get pruned. To specify the edge set of G/H we define $\phi : V(G) \to V(G/H)$ to be the following map:

$$\phi(x) := \begin{cases} x & x \in V(G) \setminus V(H) \\ v_H & x \in V(H) \end{cases}$$

We define the edge set of G/H to be the following set:

$$E(G/H) = \{\{\phi(x), \phi(y)\} : \{x, y\} \in E \setminus H\}.$$

Theorem 3.3.1 (Theorem 5.7 [1])**.** *If G is a non-homogeneous graph with homogeneous core H and G/H is the graph resulting from shrinking H to a one node and eliminating any*

self-loops. Then, a serial rule holds

$$\mathrm{MEO}(\Gamma_G) = \mathrm{MEO}(\Gamma_H) + \mathrm{MEO}(\Gamma_{G/H}).$$

Moreover, an optimal pmf for $\mathrm{MEO}(\Gamma_G)$ *can be constructed by coupling any two optimal pmfs for* $\mathrm{MEO}(\Gamma_H)$ *and* $\mathrm{MEO}(\Gamma_{G/H})$ *respectively.*

One may deflate the resulting quotient graph as well so as to decompose it into a homogeneous core and a quotient graph. Iterating this process allows us to split the graph into homogeneous components until we are left with a single vertex. The resulting deflation sequence gives rise to a more general serial rule, where one can couple the measures on each homogeneous core to construct an optimal measure for the original graph.

3.4 Homogeneous reducible graphs

As we have seen in section 3.3, homogeneous graphs are the building blocks for constructing an optimal measure for any non-homogeneous graph. Therefore, it would be natural to investigate the construction of an optimal measure for homogeneous graphs. To accomplish this goal, we classify homogeneous graphs based on the denseness of their subgraphs. Let $G = (V, E)$ be a homogeneous graph. Then for every $H \in \mathcal{H}$, $\theta(G) \geq \theta(H)$.

Definition 3.4.1. We say that a homogeneous graph G is *reducible* if there exists a proper vertex-induced subgraph H of G such that $\theta(H) = \theta(G)$. Otherwise, we call G irreducible.

Here we show that the deflation process extends to the case of homogeneous reducible graphs.

Theorem 3.4.2. *Let* Γ_G *be the family of spanning trees on a homogeneous graph* $G = (V, E)$ *that is reducible with respect to a proper subgraph* H*. Let* G/H *be the corresponding shrunk graph, and let* Γ_H *and* $\Gamma_{G/H}$ *be the families of spanning trees on* H *and* G/H *respectively. Finally, let* φ *be the bijection between* $\psi_2(\Gamma^*)$ *and* $\Gamma_{G/H}$ *as in [1, Lemma 5.6]. Then*

1. *the following serial rule holds*

$$\mathrm{MEO}(\Gamma_G) = \mathrm{MEO}(\Gamma_H) + \mathrm{MEO}(\Gamma_{G/H});$$

2. *a pmf* $\mu \in \mathcal{P}(\Gamma_G)$ *is optimal for* $\mathrm{MEO}(\Gamma_G)$ *if and only if* $\mu \in \mathcal{P}(\Gamma^*)$ *and its marginals* μ_H *and* $\mu_{G/H}$*, are optimal for* $\mathrm{MEO}(\Gamma_H)$ *and* $\mathrm{MEO}(\Gamma_{G/H})$ *respectively;*

3. *conversely, given* $\mu_H \in \mathcal{P}(\Gamma_H)$ *and* $\mu_{G/H} \in \mathcal{P}(\Gamma_{G/H})$ *that are optimal for their respective* MEO *problems,* $\mu_H \oplus (\varphi_*^{-1}\mu_{G/H})$ *is an optimal pmf in* $\mathcal{P}(\Gamma_G)$ *for* $\mathrm{MEO}(\Gamma_G)$;

4. *finally, for any pmf* $\mu \in \mathcal{P}(\Gamma^*)$ *with marginals* μ_H *and* $\mu_{G/H}$,

$$\mathbb{P}_\mu(e \in \underline{\gamma}) = \begin{cases} \mathbb{P}_{\mu_H}(e \in \gamma_H) & \text{if} \quad e \in E_H \\ \mathbb{P}_{\mu_{G/H}}(e \in \gamma_{G/H}) & \text{if} \quad e \in E_G \setminus E_H \end{cases}$$

Note that the quotient graph G/H is again homogeneous by Theorem 5.8 [1], so the deflation process stops once the quotient graph is irreducible.

The main conclusion of this paper is that homogeneous irreducible graphs are building blocks for the MEO problem on arbitrary graphs and that the MEO probelm on homogeneous irreducible graphs is solved by WUST.

3.4.1 Examples

Here we describe some examples of small homogeneous graphs. We begin with an irreducible example.

Example 3.4.3. Consider a square-with-diagonal graph, as in Figure 3.1. As was shown in [1, Example 2.2], this graph is homogeneous and the uniform pmf is not optimal. Note that it is also irreducible, since its denseness is $5/3$, which is strictly greater than the denseness of any of its proper vertex-induced subgraphs, e.g. triangles have denseness $3/2$ and trees have denseness 1.

In particular, the square-with-diagonal graph admits a unique set of weights that yield an optimal WUST. These weights can be computed by hand using Kirchhoff's theorem. Indeed, due to the symmetry, the weights on the four sides of the square will be the same, say a. Hence, the only other weight is on the diagonal, call it b. Let μ_σ be the σ-uniform measure induced by the edge-weights a and b, as in the left-side of Figure 3.1. By Kirchhoff's theorem for any edge e:

$$\mathbb{P}_{\mu_\sigma}(e \in \underline{T}) = \sigma(e) R_{\text{eff},\sigma}(e).$$

Thus, we can compute the optimal weights by equalizing the "per edge" effective resistances for the side edges of the square, with the one for the diagonal. Using the usual serial and parallel rule, we get that

$$\frac{b}{b+a} = \frac{a}{a + \frac{1}{\frac{1}{a} + \frac{1}{b + \frac{a}{2}}}}$$

Or, in terms of $r := a/b$,

$$\frac{1}{1+r} = \frac{1}{1 + \frac{1}{1 + \frac{1}{\frac{1}{r} + \frac{1}{2}}}}$$

Solving for r, we find that $r = 2/3$. Since $\mu_{t\sigma} = \mu_\sigma$ for any $t > 0$, we can choose to set $a = 2$ and $b = 3$. This computation yields the weighted square with diagonal on the right-side of Figure 3.1.

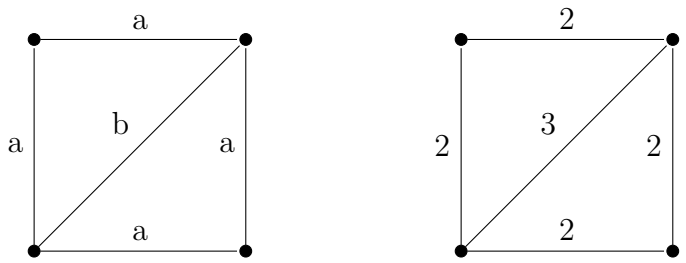


Figure 3.1: *Square with Diagonal and Optimally Weighted Square with Diagonal*

Next, we show that the House graph is homogeneous and reducible.

Example 3.4.4. Let H be the house graph as in left-side of Figure 3.2. The triangle R at the top of the House graph is called the *roof*. The two vertical *walls* in conjunction with the horizontal *pavement* form a *floor*. We consider the set of all spanning trees of H that use exactly two edges from the roof

$$\bar{\Gamma}_H = \{T \subset \Gamma_H \quad | \quad |T \cap R| = 2\}.$$

We observe that, for any $T \in \bar{\Gamma}_H$, we also have $|T \cap (H \setminus R)| = 2$, since $E(T) = V(H) - 1 = 4$. Conversely, note that, picking two edges at random from the roof and two edges at random from the floor, will always result in a spanning tree for H. Based on this observation, we conclude that $|\bar{\Gamma}_H| = 9$. Moreover, this procedure gives rise to the uniform pmf conditioned on $\bar{\Gamma}_H$:

$$\bar{\mu}(T) = \begin{cases} \frac{1}{9} & T \in \bar{\Gamma}_H \\ 0 & \text{otherwise.} \end{cases}$$

We claim that $\bar{\mu}$ is optimal for H and H is homogeneous. Moreover, H is reducible and $\bar{\Gamma}_H$ is the family of all fair trees.

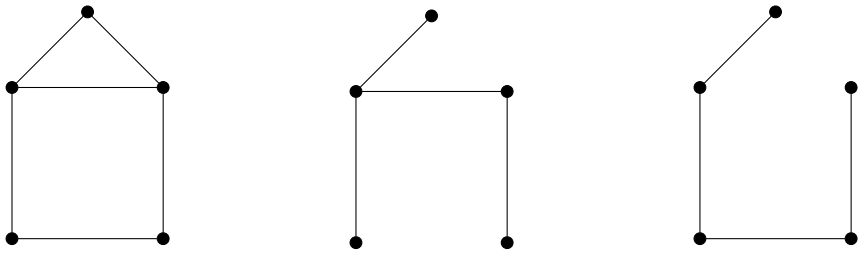


Figure 3.2: *House, Fair and Forbidden Trees from left to right*

Observe that the denseness of H is

$$\theta(H) = \frac{|E(H)|}{|V(H)| - 1} = \frac{6}{4} = \frac{3}{2}.$$

Moreover, $\theta(K) \leq \theta(H)$ for any vertex-induced subgraph K of H (note that the square has denseness $4/3$). Therefore, H is homogeneous with edge probabilities equal to $2/3$. Furthermore, the roof R also has denseness $3/2$, so H is reducible with respect to R.

Homogeneity can also be deduced by computing the edge probabilities of the pmf $\bar{\mu}$ defined above using the deflation method. Indeed, a $\bar{\mu}$-random tree can be constructed by picking two edges out of three from the roof, then deflating the roof and considering the quotient graph G/R which happens to be a triangle as well, and picking two edges at random from there. Since the edges of G/R are in one-to-one correspondence with the edges of the floor in H, we see that every $e \in E(H)$ belongs to exactly 6 out of 9 equally probable spanning trees.

In particular, this shows that $\bar{\Gamma}_H \subset \Gamma_H^f$. To see the other inclusion, assume $T \in \Gamma_H \setminus \bar{\Gamma}_H$. Then $|T \cap R| \in \{0, 1, 3\}$, however, 0 implies that one vertex is isolated in T, and 3 would yield a cycle in T. So, we must have $|T \cap R| = 1$. In particular, $T \cap (H \setminus R)| = 3$ and letting T^* be the projection of T onto the quotient graph H/R, we see that T^* is no longer a spanning tree. This means that T is forbidden, which proves that $\bar{\Gamma}_H = \Gamma_H^f$.

Next we generalize the House graph and show that there are infinite families of homogeneous reducible graphs.

Example 3.4.5. Let H_N be the N-story house, which can be constructed inductively as follows. Let $H_1 = H$, the House graph, and assume we have constructed H_{N-1}, then H_N can be obtained from H_{N-1} by appending another floor at the bottom. For convenience we number the floors in reverse order than customary in regular houses, so that the N-th floor is the last one added, namely the bottom floor. One can check that

$$\theta(H_N) = \frac{|E(H_{N-1})| + 3}{V(H_{N-1}) - 1 + 2} = \frac{3N}{2N} = \frac{3}{2}.$$

So H_N is also homogeneous and reducible with respect to the roof. Moreover, $H_N/R = H_{N-1}$. So the deflation process consists of a sequence of N reductions and the fair trees consist exactly of those that at each stage of the deflation process use exactly 2 edges from the roof.

As an aside, in this example one can compute $|\Gamma_H|$ and Γ_H^f exactly and notice that as N tends to infinite the number of forbidden trees grows much faster than the number of fair trees.

We modify this example to create an infinite family of 3-regular homogeneous reducible graphs $\{G_N\}_{N=1}^{\infty}$. Let D_N be the double-roofed N-story house, i.e., the N-story house in conjunction with a extra vertex that is connected with the two vertices on the pavement of the bottom (N-th) floor of H_N. We construct G_N by gluing D_N to D_{2N+1} by matching the tips of their two roofs, see Figure 3.3. One may observe by inspection that G_N is 3-regular.

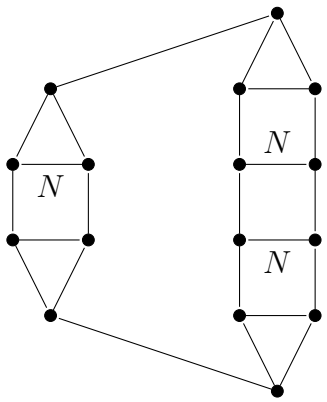


Figure 3.3: *Infinite family of reducible 3-regular graphs*

We claim that G_N is reducible with respect to D_N. Indeed,

$$\begin{aligned}\theta(G_N) &= \frac{3|E(H_N)|+6}{3|V(H_N)|} = \frac{|E(H_N)|+2}{|V(H_N)|} = \theta(D_N) \\ &= \frac{3N+5}{2N+3} = \frac{3}{2} + \frac{1}{4N+6} > \frac{3}{2}.\end{aligned}$$

Moreover, any connected vertex-induced proper subgraph K of G_N, other than D_N, must have $\theta(K) < \theta(G_N)$, so that G_N is homogeneous. To see this, think of $V(K)$ as a cut and let $S := \partial V(K) \subset E$ be its boundary. Since K is vertex-induced, $|S| > 1$, and since K is proper, $|V(G_N) \setminus V(K)| > 0$. Since G_N is planar, we can think of S as a union of cycles in the dual graph (which is obtained by putting a node in each complementary tile and adding

a link whenever two tiles share an edge). Note that there is only one pair of tiles that share more than one edge, namely the unbounded tile and the tile between D_N and D_{2N+1}. Thus, there is only one cycle of length 2 in the dual and in this case $K = D_{2N+1}$ (since $K \neq D_N$). However, one can compute the denseness of D_{2N+1} as

$$\theta(D_{2N+1}) = \frac{6N+8}{4N+5} = \frac{3}{2} + \frac{1}{8N+10} < \frac{3}{2} + \frac{1}{4N+6} = \theta(G_N).$$

In other words, unless S consists of exactly the two edges connecting D_N to D_{2N+1}, we necessarily have $|S| \geq 3$. Also, since every edge in S reduces the degree of some vertex in K by one, we have

$$2|E(K)| \leq 3|V(K)| - |S| \leq 3(|V(K)| - 1).$$

Thus,

$$\theta(K) \leq \frac{3}{2} < \theta(G_N).$$

This shows that G_N is homogeneous.

Next we show how to generalize the square-with-diagonal example and get an infinite family of homogeneous irreducible graphs.

Example 3.4.6. Let T_N be N triangles sharing an edge as in Figure 3.4. In particular, T_2 is the square with diagonal. There are 2^N spanning trees that contain the shared edge e' and there are $N2^{N-1}$ spanning trees that do not contain the shared edge. Indeed, if we pick a spanning tree that contains e', then we must select only one edge from each of the N triangles to avoid cycles, and to cover all vertices. On the contrary, if we choose a spanning tree that does not contain e', then we must select two edges from one of the N triangles. As a result we have $N2^{N-1}$ options in this case. Therefore, T_N has $(N+2)2^{N-1}$ spanning trees in total.

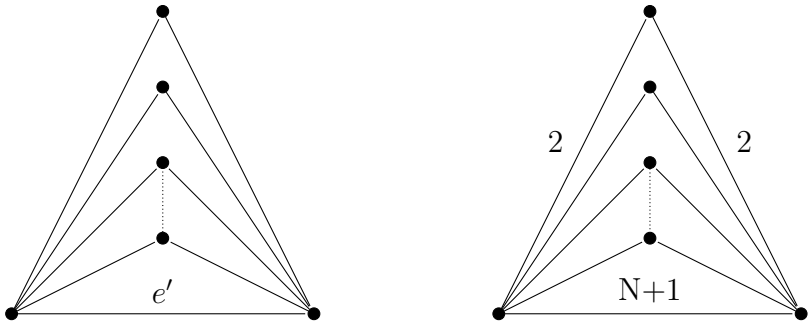


Figure 3.4: *N triangles sharing an edge*

Based on Example 3.4.3 it is natural to consider the following set of spanning trees:

$$\bar{\Gamma}_{T_N} = \{T \in \Gamma_{T_N} \quad | \quad e' \notin T\}.$$

We will see that the following pmf is optimal for T_N,

$$\bar{\mu}(T) = \begin{cases} \frac{1}{2^{N-1}(2N+1)} & T \in \bar{\Gamma}_{T_N} \\ \frac{N+1}{2^N(2N+1)} & \text{otherwise.} \end{cases}$$

Let $e \in E(T_N) \setminus \{e'\}$, then

$$\sum_{T \in \bar{\Gamma}_{T_N}} \mathcal{N}(T,e) = (N+1)2^{N-2} \qquad \text{and} \qquad \sum_{T \in \Gamma_{T_N} \setminus \bar{\Gamma}_{T_N}} \mathcal{N}(T,e) = 2^{N-1}.$$

Hence, for $e \neq e'$:

$$\mathbb{P}_{\bar{\mu}}(e \in \underline{T}) = \sum_{T \in \bar{\Gamma}_{T_N}} \bar{\mu}(T)\mathcal{N}(T,e) + \sum_{T \in \Gamma_{T_N} \setminus \bar{\Gamma}_{T_N}} \bar{\mu}(T)\mathcal{N}(T,e) = \frac{N+1}{2N+1}.$$

Since the number of trees that use the diagonal is 2^N, we have the following edge-usage probability for the diagonal:

$$\mathbb{P}_{\bar{\mu}}(e' \in \underline{T}) = \sum_{T \in \Gamma_{T_N} \setminus \bar{\Gamma}_{T_N}} \bar{\mu}(T)\mathcal{N}(T,e') = 2^N \frac{N+1}{2^N(2N+1)} = \frac{N+1}{2N+1}.$$

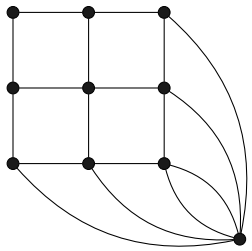


Figure 3.5: *Modified Grids*

In particular, this shows that T_N is homogeneous.

In addition, since every tree is fair, T_N is irreducible. To find the optimal weights, we equalize the per edge effective resistances for any $e \in E(T) \setminus \{e'\}$ and e'. The computations yield the optimal weights as in the right of Figure 3.4.

The graphs T_N are far from being regular. Are there infinitely many homogeneous irreducible graphs that are also 3-regular? The answer is yes, because if $G = (V, E)$ is 3-regular, and $|V| = 6n + 2$, for some $n \geq 1$, then $|E| = 9n + 3$. In particular,

$$\begin{aligned}\gcd(|E|, |V| - 1) &= \gcd(9n + 3, 6n + 1) = \gcd(3n + 2, 6n + 1) \\ &= \gcd(3n + 2, 3n - 1) = \gcd(3, 3n - 1) = 1.\end{aligned}$$

Therefore, the denseness ratio cannot be simplified and thus G is irreducible.

3.4.2 Example: modified grids

Consider a modified grid G as in Figure 3.5. It consists of a standard m-by-n grid G_0 and an extra node v_0 connected with one edge to each node on the bottom and the right hand-side of G_0. In particular, the bottom-right corner of G_0 is connected with two edges to v_0. Alternatively, G is the graph obtained from an $(m+1)$-by-$(n+1)$ by identifying every node on the bottom and right hand-sides. Note that G has $mn + 1$ vertices and $2mn$ edges. In particular, it has denseness $\theta(G) = 2$.

We associate to every vertex $v \in V(G_0)$ in the grid an edge set $E_v \subset E(G)$, consisting of

the edge to the right of v and the one below v. Consider the collection of all subgraphs of G that use exactly one edge from each E_v:

$$\bar{\Gamma}_G := \{T \subset E \quad | \quad |T \cap E_v| = 1, \quad \text{for all } v \in V(G_0)\}.$$

Notice that every such subgraph must be a spanning tree of G, so that $\bar{\Gamma}_G \subset \Gamma_G$. To see this, observe first that every $T \in \bar{\Gamma}_G$ is connected and spans $V(G)$. Indeed, starting from any node $v \in V(G_0)$, one can follow an edge and either move to the right or down, until eventually reaching v_0. Also, any such T satisfies

$$|E(T)| = \sum_{v \in V(G_0)} |E_v \cap T| = \sum_{v \in V(G_0)} 1 = |V(T)| - 1. \tag{3.4.1}$$

Let $\bar{\mu}$ be the uniform pmf conditioned on $\bar{\Gamma}_G$:

$$\bar{\mu}(T) := \begin{cases} \frac{1}{|\bar{\Gamma}_G|} & \text{if } T \in \bar{\Gamma}_G \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $\bar{\Gamma}_G$ or $\bar{\mu}$ can be thought as the random spanning trees generated by tossing independently a fair coin at each node of the standard grid and either moving to the right or downward. In particular, notice that every tree $T \in \bar{\Gamma}_G$ has a *partner tree*, i.e., the tree that can be obtained by substituting every edge in $T \cap E_v$ with the edge $E_v \setminus T$.

We claim that Γ_G *this gives rise to the family of all fair trees for the modified grid, i.e.,* $\Gamma_G^f = \bar{\Gamma}_G$*, and that* $\bar{\mu}$ *is optimal for* MEO(Γ_G).

Moreover, we will see that $\bar{\Gamma}$ consists exactly of all the trees $\gamma \in \Gamma_G$ that have the *partner tree property*, i.e., that the complement $E \setminus \gamma$ is also a tree.

To see that $\bar{\Gamma} \subset \Gamma_G^f$, it is enough to show that $\bar{\mu}$ is optimal for MEO(Γ_G). Notice that every edge $e \in E(T)$ lies in $T \cap E_v$ for some vertex $v \in V(G_0)$. Since only half of the trees of

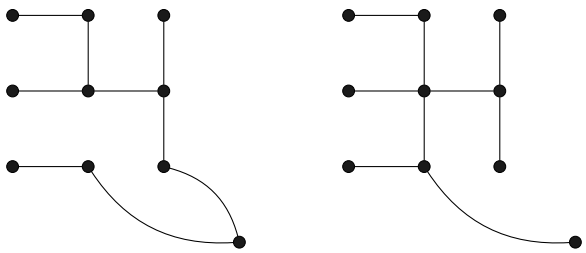


Figure 3.6: *Fair tree on the left and Forbidden tree on the right*

$\bar{\Gamma}_G$ use e by construction, we will have the following result for the edge-usage probabilities

$$\mathbb{P}_{\bar{\mu}}(e \in \underline{T}) = \sum_{T \in \bar{\Gamma}_G} \bar{\mu}(T)\mathcal{N}(T, e) = \frac{1}{|\bar{\Gamma}_G|} \sum_{T \in \bar{\Gamma}_G} \mathcal{N}(T, e) = \frac{1}{2},$$

Coincidentally, this also proves that G is homogeneous.

Before proving the other inclusion, we examine the reducibility of G. Let H be the subgraph of G induced by the special vertex v_0 and the corner of the standard grid, which thus has a double edge. Then G is reducible with respect to H, because H also has denseness equal to 2. If we deflate G with respect to H, the resulting quotient graph G/H is a standard grid such that the bottom right corner has additional edges connecting to each node to its left and each node above it. In particular, since we the edge count decreased by 2 and the node count by 1, G/H still has denseness equal to 2. Also, we created additional 2-vertices and 2-edges subgraphs, and we can thus repeat the deflation process. We can, for instance, continue applying the deflation process going from right to left, row by row, until we are left with a 2-vertices and 2-edges graph. Moreover, observe that at each step the two edges that get removed are exactly the pairs of edges in one of the sets E_v for $v \in G(V_0)$ defined above.

Now assume that $T \in \Gamma_G \setminus \bar{\Gamma}_G$. Then, by definition of $\bar{\Gamma}$, and the Pigeonhole principle, there exists vertices $v^*, v_* \in V(T)$, such that $|T \cap E_{v^*}| = 2$ and $|T \cap E_{v_*}| = 0$. Indeed, T must choose mn edges from E, which is partitioned into mn pairs of edges E_v. See Figure 3.6. However, this means that in the deflation process described above, we are guaranteed to get to a point where the two edges in, say, E_{v_*} connect v_* to a single other node u_*, so as to form a subgraph H_* of denseness two in a larger homogeneous graph G_* of the same

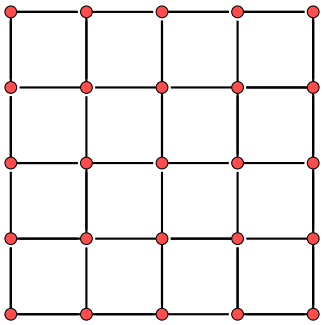


Figure 3.7: *Grid*

denseness. Let T_* be the tree obtained from T by removing all the pairs E_v that have been involved in the deflation thus far. Then, T_* does not restrict to a spanning tree of H_*, and thus T^* is forbidden.

3.4.3 Example: grids

Let $G_{m,n}$ be the m-by-n planar grid graph, in the sense that

$$V = \big\{(i,j) \in \mathbb{Z}^2 : 0 \le i < m \text{ and } 0 \le j < n\big\}$$

and for $0 \le i < m$ and $0 \le j < n$,

- $\{(i,j),(i,j+1)\} \in E \quad \text{iff} \quad j+1 < n$
- $\{(i,j),(i,j-1)\} \in E \quad \text{iff} \quad j-1 \ge 0$
- $\{(i,j),(i+1,j)\} \in E \quad \text{iff} \quad i+1 < m$
- $\{(i,j),(i-1,j)\} \in E \quad \text{iff} \quad i-1 \ge 0$

Recall that for a graph $G = (V,E)$ with no self-loops, the *denseness* of G is

$$\theta(G) := \frac{|E|}{|V|-1}$$

Moreover, if the denseness of G is strictly greater than the denseness of any of its proper vertex-induced subgraphs H, then we say that G is *irreducible*.

Lemma 3.4.7. *The planar grid $G_{m,n}$ is irreducible.*

Proof. First note that:

$$|V(G_{m,n})| = mn \qquad \text{and} \qquad |E(G_{m,n})| = (m-1)n + m(n-1) = 2mn - m - n$$

So

$$\theta(G_{m,n}) = \frac{2mn - m - n}{mn - 1} = 2 - \frac{m + n + 2}{mn - 1}$$

In particular,

$$\frac{\partial \theta(G_{m,n})}{\partial m} = -\frac{(mn-1) - n(m+n+2)}{(mn-1)^2} = \frac{(n+1)^2}{(mn-1)^2} > 0,$$

and by symmetry the same holds for the derivative in n, meaning that the denseness of rectangular grids is strictly increasing with the size of the grid, is always less than 2, and in fact tends to 2 as $m, n \to \infty$.

Now let H be a connected vertex-induced subgraph of $G_{m,n}$. We first consider the smallest rectangular grid containing H. Namely, let

$$i_{min} := \min\{i : \ \exists j, (i,j) \in V(H)\}$$

and similarly define i_{max}, j_{min}, and j_{max}. Then, define the translated grid

$$\tilde{G} := (i_{min}, j_{min}) + G_{i_{max}-i_{min}, j_{max}-j_{min}}.$$

If $H = \tilde{G}$, then $\theta(H) = \theta(\tilde{G}) < \theta(G)$ and we are done. If H is a proper subgraph of $\tilde{G}$, then since $\theta(\tilde{G}) \leq \theta(G)$, it will be enough to show that $\theta(H) < \theta(\tilde{G})$. In particular, without loss of generality we can assume that $\tilde{G} = G_{m,n}$.

As an embedded planar graph $G_{m,n}$ has $(m-1)(n-1)$ bounded faces. Whenever such a

face has all four vertices in $V(H)$ we fill it in so that H becomes a connected compact set in the plane $\hat{H}$. In particular, the complement of $\hat{H}$ is an open set consisting of finitely many bounded components Ω_j, $j = 1, \ldots, k$, and one unbounded component Ω_0. The boundary of each one of these components can be parametrized by a curve Γ_j, $j = 0, \ldots, k$. This can be seen by thickening $\hat{H}$ a little bit, for instance by defining

$$\hat{H}_\epsilon := \bigcup_{z \in \hat{H}} \{w \in \mathbb{C} : |w - z| \leq \epsilon\}.$$

For each ϵ, the boundary of $\hat{H}_\epsilon$ consists of C^1-smooth Jordan curves $\Gamma_{j,\epsilon}$, $j = 0, \ldots, k$. If we parametrize each $\Gamma_{j,\epsilon}$ with its arc-length parametrization, then, as ϵ tends to 0, they converge uniformly to the curves Γ_j.

We begin by looking at the boundary of the unbounded component Ω_0 and think of it as being parametrized counter-clockwise. In particular, Γ_0 can be taken to be piecewise linear so that the derivative Γ_0' is well defined away from the nodes of $\mathbb{Z}^2$. Since $\Gamma_{0,\epsilon}$ bounds the simply connected domain $\Omega_{0,\epsilon} \cup \{\infty\}$, by the argument principle, the change in argument for the derivative $\Gamma_{0,\epsilon}'$, as the curve completes one full loop, is 2π. As ϵ tends to 0, this property is inherited by Γ_0, as long as the argument of the derivative Γ_0' is properly defined. At any moment when Γ_0 is not at a node of $\mathbb{Z}^2$, the right hand of a walker traveling along with Γ_0 will always be touching Ω_0. In particular, the only changes in the argument of Γ_0' that are allowed when Γ_0 passes through a node of $\mathbb{Z}^2$ are

$$0, \frac{\pi}{2}, \pi, \text{ and, } -\frac{\pi}{2}.$$

In words, either the walker goes straight, turns 90° left, does a 180° turn in the positive direction, or turns 90° right. This can be verified by looking at the unbounded component $\Omega_{0,\epsilon}$ of the thickened $\hat{H}_\epsilon$. Indeed, if Γ_0 were to do a 180° turn in the negative direction, then while walking along $\Gamma_{0,\epsilon}$ the right hand would be touching the thickened neighborhood of an edge, and this neighborhood would disappear as ϵ tends to 0, so that edge would not be part of the boundary of Ω_0, which leads to a contradiction.

Now assume $v \in \mathbb{Z}^2$ is a node where Γ_0 makes a right turn, and let $u \in \mathbb{Z}^2$ be the node visited by Γ_0 just before v, and $w \in \mathbb{Z}^2$ the one visited just after v. Then, u, v, w bound a face f of $G_{m,n}$ and the fourth corner t does not belong to H. To see why, assume by contradiction that all four corners are in $V(H)$, then the face would be contained in $\hat{H}$. But then the right hand would not be touching the unbounded component Ω_0 in this case. Also, note that since H is vertex-induced the two sides of f incident at t also do not belong to $E(H)$.

Now, if we add t to $V(H)$, to get a new vertex-induced graph H'. Since one of the coordinates of t is equal to one of the coordinates of either u or w, we see that H' is still a subgraph of $G_{m,n}$. Moreover, we are guaranteed that

$$E(H') \geq E(H) + 2.$$

In particular,

$$\theta(H') = \frac{E(H')}{V(H') - 1} \geq \frac{E(H) + 2}{V(H)}$$

and

$$\begin{aligned} \frac{E(H) + 2}{V(H)} > \frac{E(H)}{V(H) - 1} &\iff (E(H) + 2)(V(H) - 1) > E(H)V(H) \\ &\iff 2V(H) > E(H) \\ &\iff 4V(H) > 2E(H) = \sum_{x \in V(H)} \deg_H(x) \end{aligned}$$

where the last equality follows from the Handshake Lemma. However, since H is a subgraph of the grid $G_{m,n}$, we always have $\deg_H(x) \leq 4$, and moreover, at least one vertex of H has $\deg_H(x) < 4$. This proves that $\theta(H') > \theta(H)$.

If $H' = G_{m,n}$, we are done. Otherwise, we can repeat the same argument with H replaced by H'. This process has to end, so without loss of generality, we can assume that Γ_0 never makes any right turns. In particular, the argument of Γ_0' can only change by 0, $\pi/2$, or π. So since there have to be an even number of argument changes that are positive, and the

sum must be 2π, we either get two changes by π or four changes by $\pi/2$. In either case, we see that Γ_0 describes the boundary of a rectangular grid, and by assumption this grid has to be $G_{m,n}$.

Finally, consider a bounded component Ω_j with $j \geq 1$. Assume that the boundary of Ω_j is described in the clockwise direction by a curve Γ_j, constructed as Γ_0 above, using the thickened set $\hat{H}_\epsilon$. Once again the allowed argument changes at a grid node are $0, \pi/2, \pi$, and $-\pi/2$. Also, when walking along Γ_j the right-hand touches Ω_j and the total argument change must equal -2π. In particular, there must always be at least four right turns with argument change $-\pi/2$. By repeating the argument above we see that each right turn identifies a vertex in Ω_j that can be added to H in a way to increase the denseness, and therefore after finitely many steps we get that the component Ω_j has been filled in. Hence, after finitely many steps, $H = G_{m,n}$. This proves the lemma. □

Chapter 4

Modulus and maximum entropy

In this chapter we introduce the notion of modulus of families of objects. Then we provide a summary of results connecting modulus problem with MEO problem. In addition, we briefly discuss a spanning tree modulus algorithm introduced in [1]. In fact, spanning tree modulus is the key ingredient for solving the MEO problem. It turns out, to solve the MEO problem, it is enough to consider the maximum entropy problem over the set of optimal pmfs for modulus problem.

Finally, we present results from [2], which prove that every homogeneous irreducible graph admits a weighted uniform spanning tree pmf.

4.1 Modulus of families of objects

Assume $G = (V, E, \sigma)$ is a weighted graph with edge weights $\sigma \in \mathbb{R}^E_{>0}$. Let Γ be a countable index set. For instance, Γ can be a family of spanning trees, paths or cuts on G. Let $\mathcal{N} \in \mathbb{R}^{\Gamma \times E}_{\geq 0}$ be the *usage matrix* for Γ, i.e., each object $\gamma \in \Gamma$ is in correspondence with a usage vector $\mathcal{N}(\gamma, .)^T \in \mathbb{R}^E_{\geq 0}$, where $\mathcal{N}(\gamma, e)$= the usage of e by γ. Let $\rho : E \to \mathbb{R}_{\geq 0}$, i.e., $\rho(e)$ = the cost of using edge e, for every $e \in E$. The *total usage cost* for a given ρ, for an

object γ, is the following quantity:

$$\ell_\rho(\gamma) := \sum_{e \in E} \mathcal{N}(\gamma, e)\rho(e) = (\mathcal{N}\rho)(e).$$

We say that a density is *admissible* for a family Γ if $\ell_\rho(\gamma) \geq 1$ for every $\gamma \in \Gamma$. In other words, ρ is admissible for Γ, if "everyone pays at least a dollar". In the future we will be using the usage matrix notation while referring to admissibility, namely,

$$\mathcal{N}\rho \geq \mathbf{1},$$

where $\mathbf{1}$ is the vector of all ones in $\mathbb{R}^\Gamma$. Let Adm Γ be the set of all admissible densities in $\rho-$space:

$$\text{Adm } \Gamma := \{\rho \in \mathbb{R}^E_{\geq 0} : \mathcal{N}\rho \geq \mathbf{1}\}.$$

This set is convex, closed and has the property that adding a non-negative vector z to an admissible density ρ does not affect admissibility. More formally,

$$\text{Adm } \Gamma + \mathbb{R}^E_{\geq 0} = \text{Adm } \Gamma.$$

We define the *Fulkerson dual* of Γ is the following set:

$$\hat{\Gamma} := \text{ext}(\text{Adm } \Gamma) \subset \mathbb{R}^E_{\geq 0},$$

where ext(A) denotes the set of extreme points of A. Since $\hat{\Gamma}$ is a set of points in ρ-space, we can interpret it as a dual family of objects Γ, which, too, has its own dual family. It is known that the objects in $\hat{\Gamma}$ are identical with their usage vectors.

For fixed $1 \leq p < \infty$, the $p-$*modulus* of a family of objects Γ is:

$$\text{Mod}_{p,\sigma}(\Gamma) = \inf_{\mathcal{N}\rho \geq 1} \sum_{e \in E} \sigma(e)\rho(e)^p.$$

We define the *energy* of the density ρ to be the following quantity:

$$\mathcal{E}_{p,\sigma}(\rho) := \sum_{e \in E} \sigma(e)\rho(e)^p.$$

One may interpret the modulus problem geometrically as computing the $p-$norm distance from the convex set Adm Γ to the origin in $\mathbb{R}^E$. As a result, strict convexity of the $p-$norm guarantees the existence of a unique optimal density ρ^*, for $1 < p < \infty$. In the special case when $p = 2$ Fulkerson duality for modulus implies the following:

$$\mathrm{Mod}_{2,\sigma}(\Gamma)\mathrm{Mod}_{2,\sigma^{-1}}(\hat{\Gamma}) = 1.$$

Moreover, the extremal density η^* for $\mathrm{Mod}_{2,\sigma^{-1}}(\hat{\Gamma})$ and the extremal density ρ^* for $\mathrm{Mod}_{2,\sigma}(\Gamma)$ are connected by the following formula:

$$\eta^*(e) = \frac{\sigma(e)}{\mathrm{Mod}_{2,\sigma}(\Gamma)}\rho^*(e).$$

Let $G = (V, E)$ be a graph and let $\Gamma = \Gamma_G$ be the family of all spanning trees of G. In this case, the Fulkerson dual family $\hat{\Gamma}$ is the set of the (weighted) feasible partitions [13].

Definition 4.1.1. A *feasible partition* P of a graph $G = (V, E)$ is a partition of the vertex set V into two or more subsets, $\{V_1, ..., V_{k_P}\}$, such that each of the induced subgraphs $G(V_i)$ is connected. The corresponding edge set, E_P, is defined to be the set of edges in G that connect vertices belonging to different V_i's.

The Fulkerson dual of Γ_G is the set of all vectors

$$\frac{1}{k_P - 1}\mathbb{1}_{E_P},$$

with P ranging over all feasible partitions. Also, by Fulkerson duality the extreme points of $\mathrm{Adm}(\hat{\Gamma})$ are spanning trees. Convexity implies that any $\mu \in \mathcal{P}(\Gamma)$ induces an admissible density $\eta = \mathcal{N}^T\mu \in$ Adm $\hat{\Gamma}$. In particular, the unique optimal density η^* for $\mathrm{Mod}_2(\hat{\Gamma})$ belongs

to the convex hull of Γ. More formally, there exists an optimal pmf $\mu^* \in \mathcal{P}(\Gamma)$ such that

$$\eta^*(e) = \sum_{\gamma \in \Gamma} \mu^*(\gamma)\mathcal{N}(\gamma, e) = (\mathcal{N}^T\mu^*)(e) = \mathbb{E}_{\mu^*}(\mathcal{N}(\underline{\gamma}, e)) \quad \forall e \in E.$$

In Appendix A we introduce KKT conditions and Lagrangian duality which give rise to the following theorem.

Theorem 4.1.2 ([3]). *Let $G = (V, E)$ be an unweighted graph and $\Gamma = \Gamma_G$ be the family of spanning trees of G, and let $\hat{\Gamma}$ be its Fulkerson dual family. Then $\rho \in \mathbb{R}^E_{\geq 0}, \eta \in \mathbb{R}^E_{\geq 0}$ and $\mu \in \mathcal{P}(\Gamma)$ are optimal respectively for $\mathrm{Mod}_2(\Gamma), \mathrm{Mod}_2(\hat{\Gamma})$, and $\mathrm{MEO}(\Gamma)$ if and only if the following conditions are satisfied*

$$\begin{aligned} &\rho \in \mathrm{Adm}(\Gamma), \quad \eta = \mathcal{N}^T\mu, \\ &\eta(e) = \frac{\rho(e)}{\mathrm{Mod}_2(\Gamma)} \quad \forall e \in E \\ &\mu(\gamma)(1 - l_\rho(\gamma)) = 0 \quad \forall \gamma \in \Gamma. \end{aligned}$$

Moreover,

$$\mathrm{MEO}(\Gamma) = \mathrm{Mod}_2(\hat{\Gamma}) = \mathrm{Mod}_2(\Gamma)^{-1}.$$

4.1.1 Spanning tree modulus algorithm

As we have already discussed in chapter 2, Γ_G can be very large. In this context, since every spanning tree of G gives rise to a constraint in the modulus problem, it would be computationally challenging to work with so many constraints. However, the following algorithm has been introduced in [1] which allows one to solve the problem to within a given tolerance by iteratively "growing" a subfamily $\Gamma' \subset \Gamma_G$ with approximately the same modulus.

Remark 4.1.3. The basic Algorithm 4.1.1 has been shown to converge [1] and can compute spanning tree modulus on graphs with hundreds thousands of edges in a reasonable amount of time (minutes). However, there is no known estimate for the rate of convergence. There is a different algorithm, called the Plus One algorithm [14], that can be shown to compute

Algorithm 1 *Basic p -modulus algorithm with tolerance* $\epsilon_{tol} \geq 0$

1: **given** $\Gamma' = \emptyset$ and $\rho \equiv 0$.
2: **repeat**
3: Find $\gamma \in \Gamma_G \setminus \Gamma'$ such that $l_\rho(\gamma) < 1 - \epsilon_{\text{tol}}$. Stop if none found.
4: Add γ to Γ'.
5: Optimize ρ so that $\varepsilon_p(\rho) = \text{Mod}_p(\Gamma')$.
6: **until** stopping criterion is satisfied.

modulus in polynomial time.

4.2 Maximum entropy for optimal MEO pmfs

In this chapter we will see that homogeneous irreducible graphs admit optimal weights σ such that μ_σ is optimal for MEO problem. As we aim to generate the optimal weights numerically, we would like to rely on conventional optimization techniques that would enable fast computation of weights. First we inherited the potential theory approach (this entails representing the energy as a sum of squares of the product of effective resistances and edge-weights) to solving the optimization problem, but we concluded that the objective function of MEO problem is non-convex. To resolve this issue, in section 4.2.1 we introduce the maximum entropy problem over the set of optimal pmfs for the modulus problem. We consider the dual of the objective function of this problem. This allows us to use the advantage of convex optimization machinery.

The development of the idea of entropy of random variables and processes by Claude Shannon lies in the foundations of the information theory and of the modern ergodic theory. Entropy allows us to describe the long term behavior of random processes. More information about information theoretic entropy can be found in [23].

4.2.1 Maximizing entropy

Given a graph $G = (V, E)$ we define the *entropy* of $\mu \in \mathcal{P}(\Gamma)$ to be the following:

$$H(\mu) = -\sum_{\gamma \in \Gamma} \mu(\gamma) \log\, \mu(\gamma).$$

Here we think of $0 \log 0$ as 0.

Let η^* be the expected optimal edge usage for the FEU problem. We would like to discuss the following *maximum entropy pmf problem*:

$$\begin{aligned} \underset{\mu \in \mathbb{R}^\Gamma}{\text{maximize}} \quad & H(\mu) \\ \text{subject to} \quad & \mathcal{N}^T \mu = \eta^* \\ & \mu \geq 0, \\ & \mu^T \mathbf{1} = 1. \end{aligned} \tag{4.2.1}$$

We notice that this problem has a unique solution. The existence of the solution follows from the fact that we are maximizing a continuous function over a compact set, while uniqueness is a consequence of H being strictly concave. Indeed, $-x \log(x)$ is strictly concave since its derivative $-1 - \log(x)$ is strictly decreasing. The sum of strictly concave functions is strictly concave, hence $H(\mu)$ is strictly concave.

Additionally, we may define $H(\mu) = -\infty$ for $\mu \ngeq 0$. This allows us to remove the constraint $\mu \geq 0$ in (4.2.1). In the sequel, we will omit this constraint, with the understanding that H is defined globally this way.

The constraint $\mathcal{N}^T \mu = \eta^*$ implies that μ is optimal for the MEO and modulus problem, see for instance (3.1.3). Recall that Γ_F, the family of fair trees of G, consists of trees that are in the support of optimal pmfs.

Lemma 4.2.1. *The support of the optimal pmf that attains the maximum entropy is all of* Γ_F.

Proof. Let μ^* be an optimal pmf for modulus such that supp $\mu^* = \Gamma_F$. Then we assume

there exists an optimal pmf μ for the modulus problem whose support strictly lies inside of Γ_F. Then we consider the convex combination of μ^* and $\mu : \mu_\epsilon = (1-\epsilon)\mu + \epsilon\mu^*$ for $\epsilon \in [0,1]$. By convexity of H, μ_ϵ is optimal for modulus and $\text{supp } \mu_\epsilon = \Gamma_F$ for all $\epsilon > 0$. The goal is to show that $H(\mu_\epsilon) > H(\mu)$ for sufficiently small ϵ.

$$H(\mu_\epsilon) = -\sum_{\gamma \in \Gamma_F} \mu(\gamma) \log \mu(\gamma).$$

We investigate two distinct cases conditioning on γ. For any tree $\gamma \in \Gamma_F$ such that $\mu(\gamma) > 0$

$$-\mu_\epsilon(\gamma) \log \mu_\epsilon(\gamma) = -\mu(\gamma) \log \mu(\gamma) + O(\epsilon).$$

If $\gamma \in \Gamma_F$ such that $\mu(\gamma) = 0$, then

$$-\mu_\epsilon(\gamma) \log \mu_\epsilon(\gamma) = -\epsilon\mu^*(\gamma) \log \epsilon + O(\epsilon),$$

which implies that $H(\mu_\epsilon) > H(\mu)$ when ϵ goes to 0. □

In other words, any μ whose support is strictly inside of Γ_F cannot be optimal for the maximum entropy problem.

Based on all these considerations, we can rewrite the maximum entropy problem as follow:

$$\begin{aligned} \underset{\tilde{\mu} \in \mathbb{R}^{\Gamma_F}}{\text{maximize}} \quad & H(\tilde{\mu}) \\ \text{subject to} \quad & {\mathcal{N}_F}^T \tilde{\mu} = \eta^* \\ & \tilde{\mu}^T \mathbf{1} = 1, \end{aligned} \tag{4.2.2}$$

where $\mathcal{N}_F$ is the usage matrix for Γ_F.

4.3 When all trees are fair

In this section we discuss the maximum entropy problem with the intention to restrict to the case when all trees are fair.

4.3.1 Finding weights

The main idea is that the optimal MEO pmf that maximizes entropy should be a WUST. As we will see this is false in general, but it is true if all trees are fair. To begin, we modify the definition of WUST, see Definition 2.1.1.

Definition 4.3.1. We say that μ is a *WURST* (weighted uniform restricted spanning tree), if it is the restriction of a WUST to the set of fair trees. Namely, there is a set of positive edge weights $\sigma : E \to \mathbb{R}_{>0}$ such that $\mu(\gamma)$ is proportional to $\prod_{e\in\gamma} \sigma(e)$ whenever γ is a fair tree in Γ_F, and is 0 otherwise.

Theorem 4.3.2. *If μ^* is the maximum entropy optimal pmf, then it is a WURST, i.e., there exists a set of positive edge weights $\sigma : E \to \mathbb{R}_{>0}$ such that*

$$\mu^*(\gamma) = \frac{\prod_{e\in\gamma} \sigma(e)}{\sum_{\gamma'\in\Gamma_F} \prod_{e\in\gamma'} \sigma(e)} \quad \textit{for all} \quad \gamma \in \Gamma_F. \tag{4.3.1}$$

Proof. Let $\tilde{\mu^*} \in \mathbb{R}^{\Gamma_F}$ be the restriction of μ^* to Γ_F. Then $\tilde{\mu^*}$ is the unique minimizer for problem (4.2.2). By Lemma 4.2.1, $\tilde{\mu^*}(\gamma) > 0$ for all $\gamma \in \Gamma_F$. Since the objective function $H(\tilde{\mu})$ is smooth near the minimizer, the minimizer is characterized by the KKT conditions. To find the optimal $\tilde{\mu}$ we work with the Lagrangian of problem (4.2.2),

$$\mathcal{L}(\tilde{\mu}, u, \omega) = H(\tilde{\mu}) + u^T(\mathcal{N}_F^T\tilde{\mu} - \eta^*) + \omega(\tilde{\mu}^T\mathbf{1} - 1),$$

where $u \in \mathbb{R}^E$ and $\omega \in \mathbb{R}$.

Then we may apply the stationarity condition to $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial \tilde{\mu}(\gamma)} = -\log \tilde{\mu}^*(\gamma) - 1 + \sum_{e \in \gamma} u^* + \omega^* = 0 \quad \text{for all} \quad \gamma \in \Gamma_F.$$

By solving the equation for $\tilde{\mu^*}(\gamma)$ we get that

$$\tilde{\mu^*}(\gamma) = e^{\omega^* - 1} \prod_{e \in \gamma} e^{u^*(e)} = C \prod_{e \in \gamma} \sigma(e),$$

where $C = e^{\omega^* - 1}$ and we define $\sigma(e) := e^{u^*(e)}$. Taking into consideration that $\tilde{\mu}^*$ is a probability vector, we get that

$$C = \sum_{\gamma' \in \Gamma_F} \prod_{e \in \gamma'} \sigma(e).$$

□

Lemma 4.2.1 in conjunction with Theorem 4.3.2 yield the following result.

Corollary 4.3.3. *All trees of G are fair if and only if there exists a weighted uniform pmf μ_σ that is optimal for modulus.*

Proof. If all trees are fair, then Theorem 4.3.2 yields a WURST pmf μ^* whose support is all of Γ.

Conversely, if there exists a weighted uniform pmf μ_σ that is optimal for modulus, then all trees are fair because the weights $\sigma(e)$ are positive. □

4.3.2 Dual problem

Here we consider the dual problem to (4.2.1). To begin we assume that all trees are fair, so that $\Gamma_F = \Gamma$. We would like to find the optimal weights σ in (4.3.1). To do so we discuss the following characterization for the dual objective for the maximum entropy problem.

Theorem 4.3.4. *The optimal weights σ in (4.3.1) minimize the following energy*

$$g(\sigma) = \log\left(\sum_{\gamma\in\Gamma}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}\right) - \sum_{e\in E}\eta^*(e)\log\sigma(e) = \log\left(\sum_{\gamma\in\Gamma}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)-\eta^*(e)}\right).$$

Proof. To find the dual objective we maximize the Lagrangian for the maximum entropy problem over μ. Recall that the Lagrangian $\mathcal{L}$ can be computed as follows

$$\mathcal{L}(\mu, u, \omega) = H(\mu) + u^T(\mathcal{N}\mu - \eta^*) + \omega(\mu^T\mathbf{1} - 1).$$

As before, by maximizing over μ, we get that

$$\mu(\gamma) = C\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}, \quad \text{where} \quad C = e^{\omega-1} \quad \text{and} \quad \sigma(e) = e^{u(e)}.$$

Now we will use these formulas to substitute μ in the Lagrangian and we will also change the dual variables from u and ω to $C > 0$ and $\sigma(e) > 0$, respectively.

$$\begin{aligned} H(\mu) &= -\sum_{\gamma\in\Gamma}\mu(\gamma)\log\mu(\gamma) = -C\sum_{\gamma\in\Gamma}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}\left(\log C + \sum_{e'\in E}\mathcal{N}(\gamma,e')\log\sigma(e')\right) \\ &= -C\log C\sum_{\gamma\in\Gamma}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)} - C\sum_{\gamma\in\Gamma}\left[\left(\sum_{e'\in E}\mathcal{N}(\gamma,e')\log\sigma(e')\right)\left(\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}\right)\right] \end{aligned}$$

Also, we compute $u^T\mathcal{N}\mu$ and $\omega\mu^T\mathbf{1}$.

$$u^T\mathcal{N}^T\mu = \sum_{\gamma\in\Gamma}\sum_{e\in E}\log\sigma(e)\mathcal{N}(\gamma,e)\mu(\gamma) = C\sum_{\gamma\in\Gamma}\left[\left(\sum_{e'\in E}\mathcal{N}(\gamma,e')\log\sigma(e')\right)\left(\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}\right)\right],$$

and

$$\omega\mu^T\mathbf{1} = (1+\log C)C\sum_{\gamma\in\Gamma}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}$$

Lastly, we plug in all the summands back into $\mathcal{L}$ to obtain the dual objective

$$g(\sigma, C) = C \sum_{\gamma \in \Gamma} \prod_{e \in E} \sigma(e)^{\mathcal{N}(\gamma, e)} - \sum_{e \in E} \eta^*(e) \log \sigma(e) - \log C - 1. \tag{4.3.2}$$

Since we want to minimize g, we can first minimize this function over C and substitute the optimal value of C

$$C = \frac{1}{\sum_{\gamma \in \Gamma} \prod_{e \in E} \sigma(e)^{\mathcal{N}(\gamma, e)}}.$$

As a result, we get the desired dual objective depending on σ only. □

4.3.3 A numerical algorithm

In view of Theorem 4.3.4, the goal is to find weights σ such that $g(\sigma)$ is minimized. By the matrix-tree theorem

$$\sum_{\gamma \in \Gamma} \prod_{e \in E} \sigma^{\mathcal{N}(\gamma, e)} = \frac{1}{n} \det \left(L_\sigma + \frac{1}{n} \mathbf{1}\mathbf{1}^{\mathbf{T}} \right),$$

where L_σ is the weighted Laplacian. So this is akin to minimizing the logdet of the Laplacian, which is a known problem in the literature, [41], the difference being the extra term involving η^*. In particular, this determinant can be computed by taking the product of all non-zero eigenvalues of the weighted Laplacian.

For a fixed $e \in E$ we have the following for the derivative of g with respect to $\sigma(e)$

$$\begin{aligned} \frac{\partial g}{\partial \sigma(e)} &= \frac{\sum_{\gamma \in \Gamma} \mathcal{N}(\gamma, e) \sigma(e)^{\mathcal{N}(\gamma, e) - 1} \prod_{e' \neq e} \sigma(e')^{\mathcal{N}(\gamma, e')}}{\sum_{\gamma \in \Gamma} \prod_{e' \in E} \sigma(e')^{\mathcal{N}(\gamma, e')}} - \frac{\eta^*(e)}{\sigma(e)} \\ &= \frac{1}{\sigma(e)} \frac{\sum_{\gamma \in \Gamma} \mathcal{N}(\gamma, e) \prod_{e' \in E} \sigma(e')^{\mathcal{N}(\gamma, e')}}{\sum_{\gamma \in \Gamma} \prod_{e' \in E} \sigma(e')^{\mathcal{N}(\gamma, e')}} - \frac{\eta^*(e)}{\sigma(e)} \\ &= \frac{\eta_\sigma(e) - \eta^*(e)}{\sigma(e)}, \end{aligned}$$

where η_σ is the usage with weights σ. By Kirchhoff's Theorem, one may derive η_σ by computing effective resistances. Also, the stationarity conditions yield that $\eta_\sigma = \eta^*$.

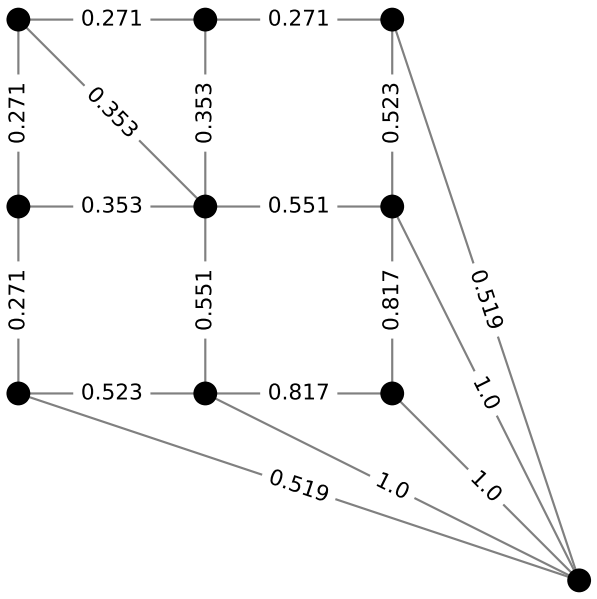


Figure 4.1: *modified grid with an extra edge*

We illustrate this with a numerical example, see Figure 4.1.

Example 4.3.5. Here we revisit Example 3.4.3 and apply the method described above to find the optimal weights for the square with diagonal. This gives an alternative way to compute these optimal weights.

First, we use the fact that all trees are fair for this graph, as we have seen in Example 3.4.3. By symmetry, we have two distinct weights on the edges of the graph. Let the weights on the sides of the square be $\beta > 0$ and the weight on the diagonal be $\alpha > 0$. Therefore, recalling that $\eta^* \equiv 3/5$, $g(\sigma)$ can be computed as follows

$$g(\sigma) = \log(4\alpha\beta^2 + 4\beta^3) - \frac{3}{5}(\log\alpha + 4\log\beta).$$

A simple calculation shows that the minimum is attained at α, β such that $\beta = \frac{2}{3}\alpha$. For instance, $\alpha = 3, \beta = 2$ minimizes g.

4.4 When not all trees are fair

We have seen in Section 4.3 that when all trees are fair there exists $\sigma(e) > 0$ for all $e \in E$ such that WURST is optimal for the maximum entropy problem. In this section we investigate the case when not all trees are fair.

4.4.1 Approximation on homogeneous graphs

In more generality, when not all trees are fair, there is no WURST that optimizes the maximum entropy problem. As we know homogeneous graphs are the building blocks for constructing an optimal pmf for any non-homogeneous graph. In this section we will provide necessary and sufficient conditions for a homogeneous graph to admit WURST. Hence, we would like to discuss the maximum entropy problem for homogeneous graphs described in Section 3.2.

Let G be a homogeneous graph. Then by definition of homogeneity every edge has the same edge-usage probability. Let κ be the reciprocal of the denseness ratio.

$$\eta^* = k := \frac{|V| - 1}{|E|}.$$

As before, we desire to solve the following dual problem with one difference only. Namely, we substituted $\eta^*(e)$ by κ for all $e \in E$.

$$\begin{aligned} &\text{minimize} \quad g(\sigma) = \log\left(\sum_{\gamma\in\Gamma_F}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}\right) - \kappa\sum_{e\in E}\log\,\sigma(e) \\ &\text{subject to} \quad \sigma > 0. \end{aligned} \tag{4.4.1}$$

First, we study some properties of $g(\sigma)$ as a function defined on $\sigma \in \mathbb{R}^E_{>0}$.

Lemma 4.4.1. *We claim that $g(\sigma)$ is one-homogeneous and is bounded from below $g(\sigma) \geq 0$.*

Proof. To show that g is one-homogeneous, we let $t > 0$ and $\sigma(e) > 0$ for all $e \in E$. Then,

we show that $g(t\sigma) = g(\sigma)$. First, we compute $g(t\sigma)$

$$g(t\sigma) = \log\left(\sum_{\gamma\in\Gamma_F}\prod_{e\in E} t^{\mathcal{N}(\gamma,e)}\sigma(e)^{\mathcal{N}(\gamma,e)}\right) - \kappa\sum_{e\in E}\log\sigma(e) - \kappa|E|\log t.$$

We may simplify the expression above by using the fact that every spanning tree contains exactly $|V|-1$ edges.

$$\begin{aligned} g(t\sigma) =& \log\left(t^{|V|-1}\sum_{\gamma\in\Gamma_F}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}\right) - \kappa\sum_{e\in E}\log\sigma(e) - \kappa|E|\log t \\ =& \log\left(\sum_{\gamma\in\Gamma_F}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}\right) - \kappa\sum_{e\in E}\log\sigma(e) + (|V|-1-\kappa|E|)\log t \\ =& \log\left(\sum_{\gamma\in\Gamma_F}\prod_{e\in E}\sigma(e)^{\mathcal{N}(\gamma,e)}\right) - \kappa\sum_{e\in E}\log\sigma(e) = g(\sigma). \end{aligned}$$

In order to show that $g(\sigma) \geq 0$, we represent $u(e)$ in terms of σ, i.e., $u(e) = \log\sigma(e)$ as we have defined $\sigma(e)$ to be $\exp u(e)$. As a result we can rewrite $g(\sigma)$ as follows:

$$g(\sigma) = \log\left(\sum_{\gamma\in\Gamma_F}\prod_{e\in E}(\exp(u(e)\mathcal{N}(\gamma,e))\right) - \log\exp\left(\sum_{e\in E}\kappa u(e)\right).$$

After applying further simplifications we get the following:

$$g(\sigma) = \log\left(\sum_{\gamma\in\Gamma_F}\prod_{e\in E}(\exp(u(e)\mathcal{N}(\gamma,e))\right) - \log\prod_{e\in E}\exp\left(\kappa u(e)\right)$$

Finally, we can represent $g(\sigma)$ as follows:

$$g(\sigma) = \log\left(\sum_{\gamma\in\Gamma_F}\prod_{e\in E}\exp\{(\mathcal{N}(\gamma,e)-\kappa)u(e)\}\right) = \log\left(\sum_{\gamma\in\Gamma_F}\exp\left\{\sum_{e\in E}(\mathcal{N}(\gamma,e)-\kappa)u(e)\right\}\right).$$

Let μ^* be an optimal pmf for the MEO problem. Since G is homogeneous, we have that the

following sum is $\kappa \sum_{e\in E} u(e)$ on average:

$$\mathbb{E}_{\mu^*}\left[\sum_{e\in E}\mathcal{N}(\underline{\gamma},e)u(e)\right] = \sum_{\gamma\in\Gamma_F}\mu^*(\gamma)\sum_{e\in E}\mathcal{N}(\gamma,e)u(e) = \sum_{e\in E}u(e)\sum_{\gamma\in\Gamma_F}N(\gamma,e)\mu^*(\gamma) = \kappa\sum_{e\in E}u(e).$$

Therefore, at least one of the trees makes the sum on the left hand side larger then the sum on the right hand side, i.e., there exists $\gamma \in \Gamma_F$ such that

$$\sum_{e\in E}(\mathcal{N}(\gamma,e)-\kappa)u(e) \geq 0,$$

which allows us to conclude that

$$g(\sigma) \geq \log(1) = 0.$$

□

Since here we have assumed that not all trees are fair, by Corollary 4.3.3, the minimizer for this problem does not exist. Hence it is natural to consider a minimizing sequence σ_k as g is bounded from below. In the lemma below we show that σ_k solves a modified version of problem (4.4.1), converges and can be normalized to 1.

Lemma 4.4.2. *The minimizing sequence has the following properties.*

1. *For all k, $\|\sigma_k\|_\infty = 1$,*

2. *For all k, σ_k is a minimizer for the problem*

$$\begin{aligned} &\underset{\sigma\in\mathbb{R}^E_{>0}}{\textit{minimize}} && g(\sigma) \\ &\textit{subject to} && \min_{e\in E}\sigma_k(e) \leq \sigma \leq 1, \end{aligned} \tag{4.4.2}$$

3. $\lim_{k\to\infty}\sigma_k = \sigma_\infty \in [0,1]^E$.

Proof. As we know g is one-homogeneous and positive, hence we may normalize σ_k by replacing it by $\sigma_k / \|\sigma_k\|_\infty$. Thus, 1. can be taken care of. Fix k and replace σ_k with the minimizer of the bound- constrained problem. First, we notice that such a replacement does not affect the fact that σ_k is a minimizing sequence. To make sure that this change does not affect (1) as well, we consider the partial derivatives of g. As we know η_{σ_k} is not equal to κ. However, the average of η_{σ_k} is κ. This fact allows us to conclude that there exists $e \in E$ such that $\eta_{\sigma_k}(e) < \kappa$. For that edge $e \in E$, we have that $\sigma_k(e) \leq 1$.

Finally, by the compactness of the unit ball in $\mathbb{R}^E$ we can extract a convergent subsequence, which guarantees (3). □

Now when we have established the existence of the limit of the minimizing sequence, we would like to study its bounds.

Lemma 4.4.3. *The limit of the minimizing sequence satisfies the following conditions*

$$\max_{e \in E} \sigma_\infty(e) = 1, \quad \min_{e \in E} \sigma_\infty(e) = 0.$$

Proof. To see the maximum property, it is enough to use continuity

$$\max_{e \in E} \sigma_\infty(e) = \lim_{k \to \infty} \|\sigma_k\|_\infty = 1.$$

For the minimum, we assume $\min_{e \in E} \sigma_\infty(e) > 0$. Therefore, σ_∞ would minimize $g(\sigma)$ over all positive σ. By stationarity μ_{σ_∞} would be optimal for the MEO problem. However, since not all trees are fair, there is no such optimizer which leads us to contradiction. □

We introduce the sets of edges with limit edge weights 0 and with positive limiting edge weights, i.e.,

$$E_0 := \{e \in E : \sigma_\infty(e) = 0\} \quad \text{and} \quad E_+ := \{e \in E : \sigma_\infty(e) > 0\}.$$

In addition, we introduce the following family of spanning trees

$$\Gamma^+ := \{\gamma \in \Gamma : |\gamma \cap E_+| = \max_{\gamma' \in \Gamma} |\gamma' \cap E_+|\}. \tag{4.4.3}$$

In other words, Γ^+ is the subfamily of spanning trees that restrict as spanning trees on each nontrivial connected component of E_+.

Example 4.4.4. Let G be the House graph as in Figure 3.2. Let $\sigma_k(e) = 1$ if e belongs to the roof, and $\sigma_k(e) = \frac{1}{k}$ otherwise. One may easily observe that

$$\sigma_\infty(e) = \begin{cases} 1, & e \in \text{roof} \\ 0, & \text{otherwise.} \end{cases}$$

Basically, E_+ is the set of edges on the roof, and E_0 consists of edges on the walls and the pavement of the House.

In addition, we notice that $\max_{\gamma' \in \Gamma} |\gamma' \cap E_+| = 2$. Therefore, Γ^+ is the family of spanning trees that contain two edges from the roof of the House. Recall that these are the fair trees for the House. Hence, Γ^+ is the family of fair trees for the House.

Next, we show that the pmf induced by this minimizing sequence allows us to concentrate the weights on trees in Γ^+.

Lemma 4.4.5. *Trees that do not restrict as spanning trees on each nontrivial connected component of E_+ for a minimizing sequence σ_k, have the following behavior in limit*

$$\lim_{k \to \infty} \mu_{\sigma_k}(\Gamma \setminus \Gamma^+) = 0.$$

Proof. Let τ be the minimum value that the limiting edge weight vector attains on E_+

$$\tau := \min_{e \in E_+} \sigma_\infty(e) > 0.$$

Let $\gamma \in \Gamma_G \setminus \Gamma^+$. Then γ restricts to a forest on some connected component of E_+. To connect

the trees of the forest, we can add an edge $e_+ \in E_+$, and to remove the cycle created in γ we should remove an edge $e_0 \in E_0$ so that the resulting graph is a tree. We denote it by γ'. Then the edge-usage of γ' can be written in terms of the edge-usage of γ

$$\mathcal{N}(\gamma', e) = \mathcal{N}(\gamma, e) + \mathbb{1}_{e^+} - \mathbb{1}_{e^-}.$$

As we have seen in Lemma 4.4.2, the minimizing sequence σ_k converges to σ_∞. More formally, for every $\varepsilon > 0$ there exists k_0 such that for $k > k_0$, $\quad \|\sigma_k - \sigma_\infty\|_\infty < \epsilon$. We will use this estimate to compare the weights of γ and γ'.

$$\prod_{e\in E} \sigma(e)^{\mathcal{N}(\gamma',e)} = \frac{\sigma_k(e_+)}{\sigma_k(e_-)} \prod_{e\in E} \sigma(e)^{\mathcal{N}(\gamma,e)} \geq \frac{\tau - \varepsilon}{\tau} \prod_{e\in E} \sigma(e)^{\mathcal{N}(\gamma,e)}.$$

The last inequity allows us to write the WURST for γ and γ' as follows.

$$\mu_{\sigma_k}(\gamma) \leq \frac{\tau}{\tau - \varepsilon} \mu_{\sigma_k}(\gamma') \leq \frac{\tau}{\tau - \varepsilon}.$$

Therefore, we get the desired result:

$$\lim_{k\to\infty} \mu_{\sigma_k}(\gamma) = 0.$$

□

4.4.2 Existence criterion for WURST

Lastly, in this section we connect the property of irreduciblity of a graph with the existence of WURST that solves the maximum entropy problem for that graph.

Let E' be a nontrivial connected component of E_+ and let V_H be the vertex set of this component. Let $H = (V_H, E_H)$ be the subgraph of G induced by the vertices V_H, i.e., V_H is the unique vertex-induced subgraph of G with the property that $E_H \cap E^+ = E'$.

Lemma 4.4.6. *Let H be the subgraph of G described as above. Then the following hold true*

1. *If* $e \in E_H \cap E_0$, *then*

$$\lim_{k\to\infty} \eta_{\sigma_k}(e) = 0.$$

2. *If* $e \in E_+$, *then for sufficiently large* k

$$\eta_{\sigma_k}(e) \leq \kappa.$$

Proof. Let $e \in E_H \cap E_0$. Then

$$\eta_{\sigma_k}(e) = \sum_{\gamma\in\Gamma\setminus\Gamma^+} \mu_{\sigma_k}(\gamma)\mathcal{N}(\gamma, e) + \sum_{\gamma\in\Gamma^+} \mu_{\sigma_k}(\gamma)\mathcal{N}(\gamma, e) = 0, \quad \text{in limit}$$

where the first sum goes to zero as k becomes large enough by Lemma 4.4.5 while the second sum is zero since $\mathcal{N}(\gamma, e) = 0$ for $\gamma \in \Gamma^+$ and for $e \in E_H \cap E_0$.

Now let $e \in E_+$. Let $\tau = \min_{e\in E^+} \sigma_\infty(e)$ and let $0 < \varepsilon << \tau$. As we have seen in Lemma 4.4.2 the limit of the minimizing sequence exists and $\|\sigma_k - \sigma_\infty\|_\infty < \varepsilon$ for $k > 0$. Therefore, for any $e \in E^+$

$$\sigma_k(e) \geq \tau - \varepsilon > \varepsilon > \min_{e'\in E} \sigma_k(e').$$

By Lemma 4.4.2 we see that the lower bound is inactive on e, therefore $\eta_{\sigma_k}(e) < \kappa$.

□

This lemma yields the following important result.

Theorem 4.4.7. *Let* H *be the vertex-induced subgraph described above. Then*

$$\frac{|E_H|}{|V_H| - 1} \geq \frac{|E_G|}{|V_G| - 1}.$$

Proof. We consider the following sum of η_{σ_k}

$$\sum_{e\in E_H} \eta_{\sigma_k}(e) = \sum_{e\in E_H} \sum_{\gamma\in\Gamma} \mu_{\sigma_k}(\gamma)\mathcal{N}(\gamma, e) \geq \sum_{e\in E_H} \sum_{\gamma\in\Gamma^+} \mu_{\sigma_k}(\gamma)\mathcal{N}(\gamma, e) = (|V_H| - 1)\mu_{\sigma_k}(\Gamma^+).$$

By Lemma 4.4.6

$$\sum_{e \in E_H} \eta_{\sigma_k}(e) \leq \kappa |E_H|.$$

Combining last two inequalities yields the following result.

$$\frac{|E_H|}{|V_H| - 1} \geq \frac{\mu_{\sigma_k}(\Gamma^+)}{\kappa} = \frac{|E_G|}{|V_G| - 1} \mu_{\sigma_k}(\Gamma^+). \tag{4.4.4}$$

Taking the limit as k approaches infinity in 4.4.4, allows us to obtain the desirable inequality.

□

Finally, we are ready to introduce the main result of this chapter.

Theorem 4.4.8. *Every homogeneous graph either admits a WURST or is reducible.*

Proof. The argument of the proof is based on the fact that either all spanning trees of G are fair or not.

If all spanning trees of G are fair, then by Corollary 4.3.3 WURST is optimal for the MEO problem.

If not all spanning trees of G are fair, then by Theorem 4.4.7 there exists a vertex-induced subgraph of G that is at least as dense as G. However, since G is homogeneous, the denisty of its subgraphs canont exceed the density of its own. Hence, G must be reducible. □

Corollary 4.4.9. *A homogeneous graph is irreducible if and only if it admits WURST.*

4.4.3 Density of WUST

Let $\mu \in \mathcal{P}(\Gamma_G)$ be a probability mass function defined on the family of spanning trees of the graph G. Let $\Gamma' \in \Gamma_G$ and let μ be uniform on Γ'. The edge-usage probabilities for every edge $e \in E$ are

$$\eta(e) = \mathcal{P}_\mu(e \in \underline{T}) = \frac{|\{\gamma \in \Gamma' : e \in \gamma\}|}{|\Gamma'|} := \frac{n(e)}{|\Gamma'|}.$$

Let $\sigma_\varepsilon = \varepsilon^{|\Gamma'| - n(e)}$.

Question 4.4.10. What can we tell about the convergence of μ_{σ_ε} as $\varepsilon \to 0$?

First we were hopeful that $\mu_{\sigma_\varepsilon} \to \mu$ as $\varepsilon \to 0$, which would allow us to have an indication that the weighted uniform pmf is dense in the set of pmfs.

Example 4.4.11. We consider the House graph and $\Gamma' = \{\gamma_1, \gamma_2\}$ - the subfamily of spanning trees γ_1 and γ_2 as in Figure 4.2. We notice that in this case both γ_1 and γ_2 are fair trees.

If $e \in \gamma$ for every $\gamma \in \Gamma'$, then $\eta(e) = 1$ and $\sigma(e) = \varepsilon^0 = 1$. If $e \notin \gamma$ for all $\gamma \in \Gamma'$, then $\sigma(e) = \varepsilon^2$. For $e \in \gamma_i$, $e \notin \gamma_j$,where $\gamma_i, \gamma_J \in \Gamma'$, $\sigma(e) = \varepsilon$.

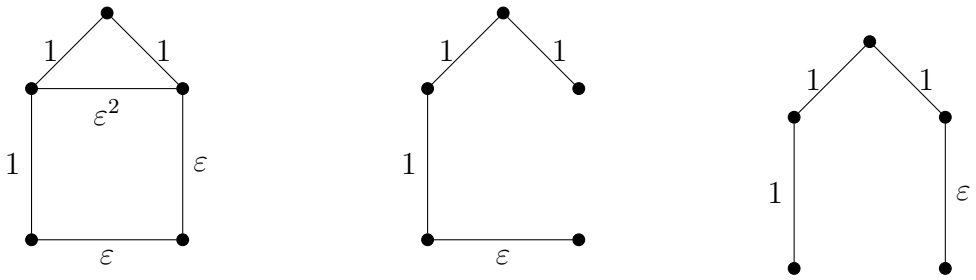


Figure 4.2: *House, γ_1 and γ_2 from left to right*

First we notice that γ_1 and γ_2 are distributed as follows

$$\mu_{\sigma_\varepsilon}(\gamma_1) = \mu_{\sigma_\varepsilon}(\gamma_2) = \frac{\varepsilon}{Z}.$$

We can compute Z by adding all the weights of spanning trees of Γ_G :

$$Z = 2\varepsilon^4 + 4\varepsilon^3 + 3\varepsilon^2 + 2\varepsilon. \tag{4.4.5}$$

Therefore, for any $\gamma \in \Gamma_G \setminus \Gamma'$

$$\mu_{\sigma_\varepsilon}(\gamma) = \frac{\varepsilon^k}{2\varepsilon^4 + 4\varepsilon^3 + 3\varepsilon^2 + 2\varepsilon},$$

where $k = 2, 3, 4$. Therefore,

$$\mu_{\sigma_\varepsilon}(\gamma) = \frac{\varepsilon^{k-1}}{2\varepsilon^3 + 4\varepsilon^2 + 3\varepsilon + 2} \to 0 \quad \text{as} \quad \varepsilon \to 0.$$

Also, the conjecture holds true in this case as

$$\mu_{\sigma_\varepsilon}(\gamma_1) = \mu_{\sigma_\varepsilon}(\gamma_2) = \frac{\varepsilon}{2\varepsilon^4 + 4\varepsilon^3 + 3\varepsilon^2 + 2\varepsilon} \to \frac{1}{2} \quad \text{as} \quad \varepsilon \to 0.$$

However, next we consider the subfamily that consists of spanning trees γ_1 and γ_2 as in 4.3. We notice that γ_1 is forbidden while γ_2 is fair. In this case we do not have the desired limit for μ_{σ_ε}.

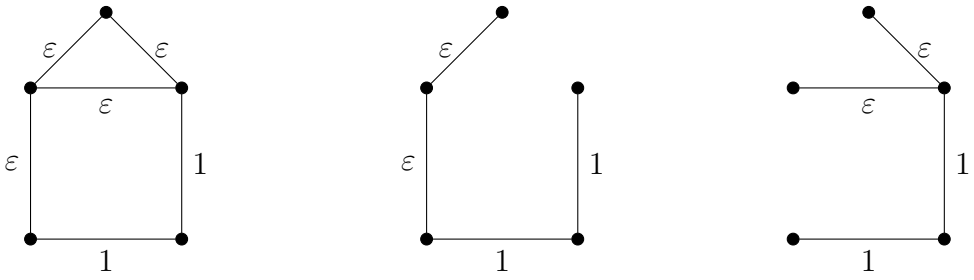


Figure 4.3: *House, γ_1 and γ_2 from left to right*

We notice that in this case the special trees are distributed as follows

$$\mu_{\sigma_\varepsilon}(\gamma_1) = \mu_{\sigma_\varepsilon}(\gamma_2) = \frac{\varepsilon^2}{6\varepsilon^3 + 5\varepsilon^2} = \frac{1}{5 + 6\varepsilon} \to \frac{1}{5} \quad \text{as} \quad \varepsilon \to 0.$$

For $\gamma \in \Gamma \setminus \Gamma'$ such that the weight of γ is proportional to ε^3 (there are 6 of them), we have the following distribution

$$\mu_{\sigma_\varepsilon}(\gamma) = \frac{\varepsilon^3}{5\varepsilon^2 + 6\varepsilon^3} \to 0 \quad \text{as} \quad \varepsilon \to 0.$$

For those trees that are not in Γ' and have weights proportional to ε^2 (there are 3 of them) we have the following distribution

$$\mu_{\sigma_\varepsilon}(\gamma) = \frac{\varepsilon^2}{5\varepsilon^2 + 6\varepsilon^3} \to \frac{1}{5} \quad \text{as} \quad \varepsilon \to 0.$$

Chapter 5

Minimization of energy

This chapter focuses on finding the optimal weights for the MEO problem for homogeneous irreducible graphs. We have seen in chapter 4 that WUST exists for this special class of graphs. In addition, we discussed the potential theory for electrical networks and provided a connection between edge-usage probabilities and effective-resistances in chapter 2. Moreover, we saw that the effective resistances can be computed via the pseudo-inverse of the Laplacian associated with a given graph.

If we assemble the information provided in previous sections, we gain an optimization problem, where the objective function is the sum of the squares of the products of effective resistances and edge-weights, over the set of all edges of the graph. This problem is non-convex. However, we show that the objective function is quasi-convex in each edge-weight. We apply rank-one update for the Laplacian of the graph to compute the optimal weights. We show that there is a unique minimizer in this case. In addition, we show that the edge-weights can be normalized to be integer-valued, which allows us to interpret them as the multiplicties of edges, thus, we are able to represent the weighted simple graph as a multigraph.

5.1 Energy of WUST

Let $G = (V, E, \sigma)$ be a weighted graph, let $\mu_\sigma \in$ WUST. In chapter 3 we discussed the MEO problem over the pmfs defined on the set of spanning trees of G. What happens if we require the optimal pmf for the MEO problem to be a WUST? In other words, we would like to solve the following minimization problem

$$\begin{aligned} &\text{minimize} && \mathbb{E}_{\mu_\sigma}|\underline{T} \cap \underline{T'}| \\ &\text{subject to} && \mu_\sigma \in \text{WUST}. \end{aligned}$$

One can simplify the objective function here in the same fashion is in chapter 3.

$$\mathbb{E}_{\mu_\sigma}|\underline{T} \cap \underline{T'}| = \sum_{e \in E} \mathbb{P}_{\mu_\sigma}(e \in \underline{T})^2.$$

Definition 5.1.1. We say that $\mathcal{E}_{\mu_\sigma}$ is the energy of μ_σ ,if

$$\mathcal{E}_{\mu_\sigma} := \sum_{e \in E} \mathcal{R}_{\text{eff}_\sigma}^2(e)\sigma^2(e).$$

The energy defined above is the objective function of the optimization problem of our interest. In addition, we can apply the potential theory introduced in chapter 2 to express the edge-usage probabilities in terms of effective resistances. Essentially, this transition allows us to express the objective function in terms of the pseudo-inverse of the Laplacian of the graph G. This gives us an opportunity to take advantage of some well-known optimization algorithms to find the optimal weights. Therefore, it makes the recovery of the corresponding WUST possible in a reasonably fast computational time. To enable the implementation of this promising road map of computations, first we use Kirchhoff's theorem, which states that

$$\mathbb{P}_{\mu_\sigma}(e \in \underline{T}) = \mathcal{R}_{\text{eff}_\sigma}^2(e)\sigma^2(e)$$

for any edge $e \in E$.

Ultimately, we desire to find σ that minimizes $\mathcal{E}(\mu_\sigma)$. For this purpose, we study the energy as a function of weights. First we substitute the effective resistance with $u^T L_\sigma^+ u$ (see chapter 2).

$$\mathcal{E}(\mu_\sigma) = \sum_{e \in E} \sigma^2(e)(u^T L_\sigma^+ u)^2,$$

where $u = \delta_y - \delta_x$ for any $\{x, y\} \in E$.

Below we show that the energy is homogeneous in weights.

Lemma 5.1.2. *The energy is homogeneous in σ, i.e., for $t > 0$,*

$$\mathcal{E}(\mu_\sigma) = \mathcal{E}(\mu_{t\sigma}).$$

Proof. We rewrite the energy in terms of the pseduo-inverse of the Laplacian L_σ^+.

$$\mathcal{E}(\mu_\sigma) = \sum_{\{x,y\} \in E} \sigma(x, y)^2 [(\delta_y - \delta_x)^T L_\sigma^+ (\delta_y - \delta_x)]^2$$

Now observe that $L_{t\sigma} = tL_\sigma$, so that $L_{t\sigma}^+ = t^{-1} L_\sigma^+$. This allows us to compute $\mathcal{E}(\mu_{t\sigma})$.

$$\begin{aligned} \mathcal{E}(\mu_{t\sigma}) &= \sum_{\{x,y\} \in E} t^2 \sigma(x, y)^2 [(\delta_y - \delta_x)^T L_{t\sigma}^+ (\delta_y - \delta_x)]^2 \\ &= \sum_{\{x,y\} \in E} t^2 \sigma(x, y)^2 [(\delta_y - \delta_x)^T t^{-1} L_\sigma^+ (\delta_y - \delta_x)]^2 = \mathcal{E}(\mu_\sigma). \end{aligned}$$

□

We would like to investigate convexity of the energy as a function of weights. However, rather than working with the energy function right away, we choose to take a detour so as to explore another problem that resembles our problem remotely. Then we move forward with the convexity exploration of our objective function.

5.2 Convexity of the total effective resistance of a graph

The total effective resistance of a graph is defined as follows

$$\mathcal{R}_{\text{tot}} := \sum_{\{x,y\}\in E} \mathcal{R}_{\text{eff}_\sigma}(x,y).$$

In [20], Boyd et al., addressed the problem of allocating a fixed total conductance among the edges so as to minimize the total effective resistance of the graph. Assume without loss of generality that the total conductance to be allocated is 1. The following optimization problem has been studied, also known as the effective resistance minimization problem

$$\begin{aligned} &\text{minimize} && \mathcal{R}_{\text{tot}} \\ &\text{subject to} && \mathbf{1}^T\sigma = 1, \quad \sigma \geq 0, \end{aligned}$$

here the inequality $\sigma \geq 0$ can be understood as a component-wise inequality.

Below we share several interpretations of the problem. In the context of electrical networks, the goal is to distribute the conductances to the branches of a circuit so as to achieve low resistance between the nodes. In the context of a Markov chain, this problem aims to select the weights on the edges so as to minimize the average hitting time between nodes. While if we think of the effective resistance between two nodes as a distance , then we can think of this as a desire to assign conductance to the graph as an attempt to make the graph small, in the sense of average distance between nodes.

It has been shown that the total effective resistance is a strictly convex function in weights. In addition, it has been provided that among all graphs with n nodes, the path has the largest value of optimal total effective resistance and the complete graph has the least.

Given a graph G, one may solve the convex optimization problem introduced above to find the optimal conductances. The total effective resistance obtained after allocating the optimal conductances, is often referred to as the *absolute total effective resistance* of the graph.

The connection between the effective resistances and the Laplacian of a graph, allows us to represent the total effective resistance problem as follows

$$\begin{aligned} \text{minimize} \quad & n \operatorname{Tr}\left(L_\sigma + \frac{\mathbf{1}\mathbf{1}^T}{n}\right)^{-1} - n \\ \text{subject to} \quad & \mathbf{1}^T\sigma = 1, \quad \sigma \geq 0. \end{aligned}$$

The Lagrange dual problem is

$$\begin{aligned} \text{maximize} \quad & h(Z, \nu, \lambda) \\ \text{subject to} \quad & \lambda \geq 0, \end{aligned}$$

where we have the following dual variables- $Z = Z^T \in \mathcal{R}^{n\times n}, \nu \in \mathcal{R}$ with the equality constraints, and $\lambda \in \mathcal{R}^m$ with the nonnegativity constraint $\sigma \geq 0$. The dual objective h is given by

$$h(Z, \lambda, \nu) = -\nu - (1/n)\mathbf{1}^T Z \mathbf{1} + 2\operatorname{Tr}(nZ)^{1/2} - n,$$

with the following constraints

$$a_i^T Z a_i \leq \nu \text{ for } i = 1, 2, ..., m, \text{ and } Z \succcurlyeq 0.$$

5.3 Quasiconvex optimization

It is known that the effective resistance is convex in weights [20]. However, despite our hope that the energy might be convex as well, it is a non-convex function. In fact, we will see in this chapter that it is quasi-convex (for definition see Appendix B) in each coordinate.

Quasi-convex optimization problems arise in various fields such as microeconomics and computer vision. It is known that quasi-convex problems can be solved by a series of convex feasibility problems [8]. In more generality solving such feasibility problems may be very costly [21]. In 1984, Nesterov [32] introduced the first efficient algorithm, namely Normalized Gradient Descent, and proved that the algorithm attains ε-optimal solution within $O(1/\varepsilon^2)$ iterations given a differentiable quasi-convex objective function. Later, in 2001, it was shown

by Kiwiel [28] that the same result holds true in case when the objective function is upper semi-continuous and quasi-convex. In addition, Konnov has shown how to attain faster rates for quasi-convex optimization [30] though it assumes that the optimal value of the objective function is known which is not the case in more generality.

5.3.1 Rank-one update approach

In this section we discuss a rank-one update approach that shows the existence of the unique minimizer for the energy. We fix an orientation on the edges. Let $e = (x, y)$ and $u_e := \delta_y - \delta_x$, where δ_i is the i-th unit vector. Let $\sigma_1 := \sigma(e_1)$. We move along the line σ_t

$$\sigma_t = \sigma + t\sigma_1 \mathbb{1}_{e_1}, \quad L_\sigma = \sum_{e \in E} \sigma(e) u_e u_e^T \quad \text{and} \quad L_{\sigma_t} = L_\sigma + t\sigma_1 u_{e_1} u_{e_1}^T.$$

As we have seen in chapter 2, the effective resistance of an edge $\{x, y\}$ can be computed by the formula 2.3.1, which allows us to compute certain entrees of the pseudo-inverse of the Laplacian associated with the weighted graph G. As a consequence, we can work with matrices instead of effective resistances. In order to make mathematical expressions succinct, we will use the following notation for effective resistances in this section:

$$R_1 := (\mathcal{R}_{\text{eff}})_\sigma(e_1) = u_{e_1}^T L_\sigma^+ u_{e_1} \quad \text{and} \quad R(e) := (\mathcal{R}_{\text{eff}})_\sigma(e) = u_e^T L_\sigma^+ u_e.$$

One may check via direct computations that the updated pseudo-inverse of the Laplacian can be computed by the following formula:

$$L_{\sigma_t}^+ = L_\sigma^+ - \frac{t\sigma_1}{1 + t\sigma_1 R_1} L_\sigma^+ u_{e_1} u_{e_1}^T L_\sigma^+,$$

which basically can also be deduced by applying Sherman-Morison formula [37].

Claim 5.3.1. *Suppose we pick an edge ab and add a weight α to it. Then the new Laplacian is*

$$L + \alpha u u^T.$$

where $u = \delta_b - \delta_a$ We claim that new pseudo-inverse is given by the formula

$$L^+ - \frac{\alpha}{1+\alpha R_1} L^+ uu^T L^+.$$

Proof. To check this it is enough to multiply the two expressions and get

$$LL^+ + \alpha uu^T L^+ - \frac{\alpha LL^+ uu^T L^+}{1+\alpha R_1} - \frac{\alpha^2 u(u^T L^+ u)u^T L^+}{1+\alpha R_1}$$

Note that in the last term $u(u^T L^+ u)u^T L^+ = R_1 uu^T L^+$. So factoring out $uu^T L^+$ on the right we get

$$LL^+ + \frac{\alpha}{1+\alpha R_1}\left[I + \alpha R_1 I - LL^+ - \alpha R_1 I\right] uu^T L^+$$

Now notice that $(I - LL^+)u = 0$.

□

Claim 5.3.2. *The energy can be computed by the following formula:*

$$\mathcal{E}_t = \frac{(1+t)^2 R_1^2}{(1+tR_1)^2} - \frac{2t}{1+tR_1} B + \frac{t^2}{(1+tR_1)^2} A + C,$$

where

$$A = \sum_{e \neq e_1} \sigma(e)^2 (u_e^T L_\sigma^+ u_{e_1})^4,\ B = \sum_{e \neq e_1} \sigma(e)^2 R(e)(u_e^T L_\sigma^+ u_{e_1})^2, C = \sum_{e \neq e_1} \sigma(e)^2 \left(u_e^T L_\sigma^+ u_e\right)^2.$$

Claim 5.3.3. *The map $t \mapsto \mathcal{E}_t$ is quasiconvex.*

The proof of these results can be found in Appendix C.

5.4 Integer valued weights

In this section, we examine the question whether the optimal weights σ can be chosen as integers, which would allow the interpretation of $\sigma(e)$ as the multiplicity of edge e. As we will be discussing iterative algorithms that converge to the minimum when the change in

energy is small enough, we would like to see whether the weights we will be updating can be normalized to be integer-valued.

Claim 5.4.1. *Let $G = (V, E)$ be a finite connected graph. Let α be the optimal weight-vector on edges. Then $\alpha = \frac{m}{n}$, where $m, n \in \mathcal{Z}$. Moreover, we can update the Laplacian as follows*

$$nL + muu^T \in \mathbb{N}.$$

Proof. We update an edge e_1 by adding α to it. Then we can compute the change in energy. The resulting function depends on α and is given by

$$\Delta\mathcal{E} = \frac{(A - 2Bc + c^2)\alpha^2 - 2(B - c^2\sigma(a,b))\alpha}{(1 + c\alpha)^2},$$

where A, B and c are the following quantities

$$\begin{aligned} A &= \sum_{\{x,y\}\in E} \sigma(x,y)^2 \left(h_{ab}(y) - h_{ab}(x)\right)^4, \\ B &= \sum_{\{x,y\}\in E} \sigma(x,y)^2 R_{\text{eff}}(x,y) \left(h_{ab}(y) - h_{ab}(x)\right)^2 \\ c &= R_{\text{eff}}(a,b). \end{aligned}$$

As a result we can compute the optimal α expressed in terms of those quantities

$$\alpha_{\text{opt}} = \frac{B - c^2\sigma(a,b)}{A - Bc + c^2 - c^3\sigma(a,b)}.$$

We desire to show that the weights are integer-valued. However, since we begin with integer weights, it is enough to show that α_{opt} is a rational number. To attain this result, we want to show that both the numerator and denominator are non-zero integers.

Let $\{u_j\}_{j=1}^N$ be the orthonormal basis of eigenvectors corresponding to the eigenvalues λ_i, $i = 1, 2, ..., N$ for the Laplacian of G. Then, as we have seen in chapter 2, we can represent

the Laplacian of G as the following sum

$$L = \sum_{j=2}^{N} \lambda_j u_j u_j^T,$$

Notice that $\lambda_1 = 0$, since G is connected and $u_1 \in < 1 >= \operatorname{Ker} L$. Again, as we have seen in chapter 2, we can write the pseudo-inverse of the Laplacian as follows

$$L^+ = \sum_{j=2}^{N} \frac{1}{\lambda_j} u_j u_j^T.$$

In addition, it is known that the pseudo-inverse of the Laplacian can be presented as follows

$$L^+ = \Big(L + \frac{J}{N}\Big)^{-1} - \frac{J}{N} = \frac{1}{N}\Big[N^2(NL + J)^{-1} - J\Big],$$

where J is an $N \times N$ matrix with 1s as entries. Recall that the inverse of a given matrix A, if it is invertible, can be computed by the following formula

$$A^{-1} = \frac{\operatorname{Adj}(A)}{\det A}.$$

As a result we can represent L^+ as follows

$$\begin{aligned} L^+ &= \frac{1}{N}\Big[\frac{N^2}{\det(NL + J)} \operatorname{Adj}(NL + J) - J\Big] \\ &= \frac{N}{\det(NL + J)} \operatorname{Adj}(NL + J) - \frac{1}{N} J. \end{aligned}$$

In order to compute thne psuedoinverse of the Laplacian we need to compute the determinant of $(NL + J)$. We notice that

$$\begin{aligned} Ju_1 &= \frac{N}{\sqrt{N}}\mathbf{1} = Nu_1, \\ (NL + J)u_1 &= NLu_1 + Nu_1 = Nu_1, \\ (NL + J)u_j &= N\lambda_j u_j, \qquad \text{for } j > 1. \end{aligned}$$

Now we see that the eigenvalues of $(NL + J)$ are the following

$$N, N\lambda_2, ..., N\lambda_N$$

and $u_1, u_2, ..., u_N$ are the eigenvectors correspondingly. Therefore,

$$\det(NL + J) = N^N \lambda_2 ... \lambda_N = N^{N+1}|\tau|.$$

Particularly, the representation of the effective resistance at (a, b) in terms of the pseudo-inverse of the Laplacian allows us to conclude that

$$c = R_{\text{eff}}(a, b) = (\delta_b - \delta_a)^T L^+ (\delta_b - \delta_a) \in \mathbb{Q}.$$

The denominator of this expression is $N \det(NL + J) = Z$. Therefore,

$$c \in \frac{\mathbb{N}}{Z}.$$

Next quantity we would like to show to be rational is the potential function.

Let $\sigma \in \mathbb{N}^E$ be the weights for G.

$$h_{ab} = L^+(\delta_b - \delta_a) \in \frac{\mathbb{N}}{N \det(NL + J)}, \quad A \in \frac{\mathbb{N}}{Z^4}, \quad B \in \frac{\mathbb{N}}{Z}.$$

Finally, we can show that the optimal weights are rational

$$\alpha_{\text{opt}} = \frac{B - c^2\sigma}{A - Bc + c^2 - c^3\sigma} = \frac{Z^4(B - c^2\sigma)}{Z^4 A - Z^4 Bc + Z^4 c^2(1 - c\sigma)} = \frac{m}{n},$$

where $m, n \in \mathbb{N}$.

Now we can let the new Laplacian to be $nL + muu^T \in \mathbb{N}$, instead of $L + \alpha uu^T$. □

5.5 Minimization algorithms for the energy

As we have seen in the previous section, the optimal update on a given edge α can be chosen to be integer-valued.

Coordinate descent algorithms solve optimization problems by successively performing approximate minimization along coordinate directions or coordinate hyperplanes. Particularly, minimization of a multivariate function can be achieved by minimizing it along one direction at a time.

Algorithm 2 *General descent method*

1: **given** a starting point $x \in \text{dom } f$
2: **repeat**
3: Determine a descent direction Δx
4: Line search. Choose a step size $t > 0$
5: Update $x := x + t\Delta x$
6: **until** stopping criterion is satisfied.

In the previous section we have introduced the computations for the optimal α. We will use this update for the algorithms below.

In the algorithm below we start with initial weights. We circle through every edge and update the weights optimally by increasing the weight on each edge one at a time. We compute the resulting energy and check if it is small enough.

Algorithm 3 *Coordinate descent method.*

1: **given** initial edge-weights $\sigma \equiv 1$
2: **repeat**
3: $\sigma(e) = \sigma(e) + \alpha_{opt}$
4: update the energy
5: **until** stopping criterion is satisfied. $\Delta\mathcal{E} \leq \epsilon$

Finally, we discuss an algorithm that searches for the best edge and updates that edge. Then we check the drop in the energy. We repeat this process of the best edge selection until the energy converges to the minimum. Here we consider the edge to be the best if it has the smallest per-edge effective resistance so as to be able to equalize per-edge effective

resistances. This choice can be justified by the following representation of α_{opt}

$$\alpha_{\text{opt}} = \frac{B - c^2\sigma(a,b)}{A - Bc + c^2 - c^3\sigma(a,b)},$$

where the numerator can be written as follows

$$B - c^2\sigma(a,b) = \sum_{\{x,y\}\in E} \sigma(x,y)(h_{ab}(y) - h_{ab}(x))^2 \left(\frac{\mathcal{R}_{\text{eff}}(x,y)}{r(x,y)} - \frac{\mathcal{R}_{\text{eff}}(a,b)}{r(a,b)} \right).$$

Algorithm 4 *Alternative Algorithm*

1: **given** initial edge-weights $\sigma \equiv 1$
2: **repeat**
3: detect the edge with $\min_e \mathcal{R}_{\text{eff}}(e)/r(e)$
4: $\sigma(e) = \sigma(e) + \alpha_{opt}$
5: update the energy
6: **until** stopping criterion is satisfied.$\Delta\mathcal{E} \leq \epsilon$

Since the energy can be graphed as in Figure 5.1 for the triangle, the coordinate descent and gradient descent algorithms converge in this case unlike in case of quasiconvex functions with multiple local minima.

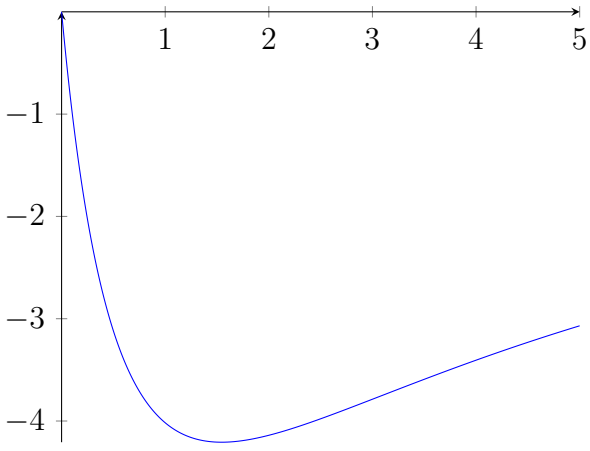


Figure 5.1: *Change in energy for the triangle*

Chapter 6

Open Problems

Alongside the questions discussed in this dissertation we would like to introduce several open questions that are currently of our interest.

6.1 The invariance of homogeneity under edge addition

Question 6.1.1. Given a homogeneous graph $G = (V, E)$ which is not a complete graph. Is there a way to add edges to it one at a time so as to maintain the homogeneity of the graph at each stage of edge addition till we recover the complete graph on $|V|$ vertices?

We consider an example of a homogeneous graph that does not preserve homogeneity with respect to edge addition if we do not choose the correct order for edge addition.

Example 6.1.2. Let $G = (V, E)$ be the modified grid as on the left in Fig.6.1 and let $G' = (V, E + e')$ be the graph obtained from G after adding a diagonal e' to the square in G as in Fig.6.1. We claim that G' is a non-homogeneous graph.

Indeed, consider the subgraph $H' \subsetneq G'$ as on the right in Fig.6.1. We notice that H' is more dense than G':

$$\theta(H') = \frac{|E(H')|}{|V(H')| - 1} = \frac{7}{3} > \frac{9}{4} = \theta(G').$$

However, if we choose to add the other diagonal to the square, we will be able to complete

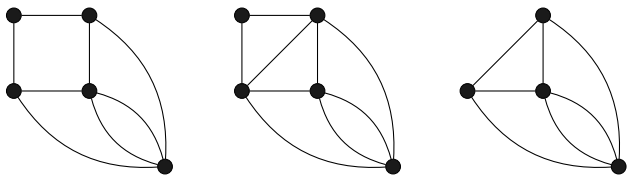


Figure 6.1: *Modified grid G, G' and its homogeneous core H' from left to right*

this graph by maintaining homogeneity. One may check this fact by computing the denseness of all vertex-induced subgraphs of this graph.

Let $G = (V, E)$ be a homogeneous graph and $G' = G + e'$ be the graph obtained from G by adding an edge e' to its edge set while leaving the vertex set intact. In other words, $G' = (V, E + e')$. Since G is homogeneous, then for every vertex-induced $H \subsetneq G$

$$\theta(H) \leq \theta(G).$$

Conjecture 6.1.3. Let $G = (V, E)$ be a homogeneous graph. Then there exists $e' \in E(K_{|V|})$ such that $G' = (V, E + e')$ is homogeneous.

Assume G' is a non-homogeneous graph. Therefore, there exists $H' \subsetneq G'$ homogeneous core, which solves the following densest subgraph problem

$$\max_{K \in \mathcal{H}} \theta(K) = \theta(H'),$$

which in conjunction with the assumption that G' is non-homogeneous implies that

$$\theta(H') > \theta(G').$$

First we notice that H' is not a subgraph of G. Indeed, if $H' \subsetneq G$, then

$$\theta(H') > \theta(G') = \frac{E+1}{V-1} > \frac{E}{V-1} = \theta(G),$$

which contradicts to G being a homogeneous graph. This argument allows us to conclude

that $e' \in E(H')$. As a result, we can present H' in terms of a proper subgraph of G ,i.e., $E(H') = E(H) + e'$ for some $H \subsetneq G$ where $V(H) = V(H')$.

Lemma 6.1.4. *If $E(H') = E(H) + e'$, then H is vertex-induced in G.*

Proof. If there exist $v_1, v_2 \in V(H)$ such that $(v_1, v_2) \in E(G) \setminus E(H)$, then $(v_1, v_2) \in E(H')$ as before since H' is vertex-induced. Therefore, $(v_1, v_2) = e' \notin E(G)$. □

If we compare the denseness of G and H then we get:

$$\theta(H) = \theta(H') - \frac{1}{V(H)-1} > \theta(G') - \frac{1}{V(H)-1} = \theta(G) + \frac{1}{V(G)-1} - \frac{1}{V(H)-1}.$$

To achieve a contradiction, it would be enough to show that there exists $H \subsetneq G$ such that $V(H) = V(H')$ where H' is the homogeneous core of G'.

$$\theta(G) - \theta(H) \geq \frac{1}{V(H)-1} - \frac{1}{V(G)-1}, \tag{6.1.1}$$

for a homogeneous graph G.

This argument leads us to the following theorem.

Theorem 6.1.5. *If G is homogeneous and reducible and $H \subsetneq G$ is a core. Then adding and edge to H so that $H' = H + e'$ is vertex-induced, will make $G' = G + e'$ heterogeneous.*

Additionally, this argument raises a question whether the existence of a vertex-induced subgraph of G that satisfies the inequality 6.1.1 guarantees that G' is homogeneous.

Conjecture 6.1.6. If G is a homogeneous graph that is not complete, then there exists $H \subsetneq G$ vertex-induced non-complete subgraph such that

$$\theta(G) - \theta(H) \geq \frac{1}{V(H)-1} - \frac{1}{V(G)-1}.$$

In the example below we apply the inequality in the conjecture to show that the modified grid admits an edge so that the resulting graph from edge addition is homogeneous.

Example 6.1.7. We consider the same example as before - the modified grid. Let G, G' and H be the graphs as in Figure 6.2.

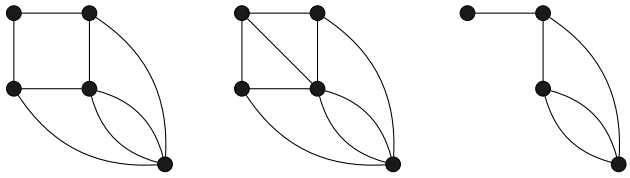


Figure 6.2: *Modified grid G, G' and H from left to right*

We see that G' is homogeneous in this case, as G satisfies the inequality 6.1.1

$$\theta(G) - \theta(H) = 2 - \frac{5}{3} = \frac{1}{3} > \frac{1}{12} = \frac{1}{3} - \frac{1}{4} = \frac{1}{|V_H| - 1} - \frac{1}{|V_G| - 1}.$$

Claim 6.1.8. *Let $G = (V, E)$ be an n-cycle ($n = |V|$). Then there exists $e' \in E(K_n) \setminus E$ such that $G' = (V, E + e')$ is homogeneous.*

Proof. Let e' be the edge from the edge-set of the complete graph on n vertices that splits G into two subgraphs so that they share e'. We choose e' so that the size of the vertex-set of the smallest subgraph of the graph is equal to $(n+1)/2$ if n is odd and to $(n+2)/2$ when n is even. We claim that this choice of e' guarantees homogeneity of G'.

Computations show that

$$\theta(G') = \frac{n+1}{n-1}.$$

Let H_1' be the subgraph of G' that has the required number of vertices and let H_2' be the other subgraph. Notice that both of these subgraphs are cycles. Hence when n is odd we have the following

$$\theta(H_1') = \frac{(n+1)/2}{(n+1)/2 - 1} = \frac{n+1}{n-1},$$

which indicates that if G' is homogeneous, then in this case it must be reducible. To convince ourselves that it is homogeneous, we also present the computations for the density of H_2'.

First, we notice that the number of vertices of H_2' is

$$|V(H_2')| = 2 + n - (n+1)/2 = \frac{n+3}{2}.$$

Plugging this back into the density formula we get the following

$$\theta(H_2') = \frac{(n+3)/2}{(n+3)/2 - 1} = \frac{n+3}{n+1} < \frac{n+1}{n-1}.$$

Similarly, if n is even we get that

$$\theta(H_1') = \theta(H_2') = \frac{(n+2)/2}{(n+2)/2 - 1} = \frac{n+2}{n} < \frac{n+1}{n-1},$$

which illustrates that G' is homogeneous, irreducible in this case. □

Example 6.1.9. In this example we start out with the 6-cycle graph as in figure 6.3.

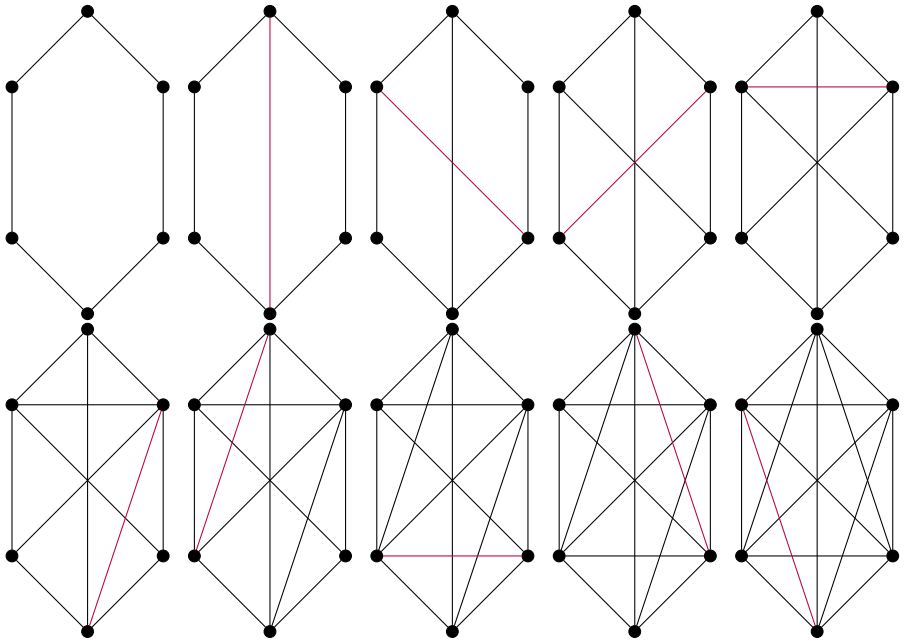


Figure 6.3: *Homogeneous completion for 6-cycle*

We would like to add edges to the 6-cycle until we obtain the complete graph on 6 vertices so that at each stage of edge addition the resulting graph is homogeneous. To complete the cycle, first we create a 3-regular graph from a cycle by constructing all $(|V|/2+1) = 4$-cycles. There are $|V|/2 = 3$ such cycles. Notice that this graph is homogeneous as it is regular. Then we create a 5-regular graph by constructing all 3-cycles. There are $|V| = 6$ such cycles.

6.2 Irreducible random regular graphs

In the Erdös-Rényi random graph $G_{n,p}$, each pair of vertices is connected by an edge with probability p. More commonly, behind the study of random graphs lies a desire to understand the properties of 'typical' graphs. Particularly, in this section we will be concerned with random regular graphs. In the context of homogeneity of a graph, it is known that almost every d-regular graph is homogeneous. The question that we would like to address here is precisely the following

Question 6.2.1. What portion of d-regular graphs on $|V|$ verices are irreducible?

We recall that irreducibility is a property of a homogeneous graph that prevents it from having subgraphs that are as dense as the graph itself. Moreover, we have the following proposition which allows us to be certain about the graphs that are irreducible if they satisfy a certain criterion.

Proposition 6.2.2. Let $G = (V, E)$ be a homogeneous graph. If $|V| - 1$ and $|E|$ are relatively prime, then G is irreducible/

Example 6.2.3. Here we discuss an example of a graph, where the condition $(|V|-1, |E|) = 1$ is violated, yet the graph turns out to be irreducible. Let $G = (10, 15)$ be the Peterson graph as in Figure 6.4. Notice that the density for the Petersen graph is $5/3$,

$$\theta(G) = \frac{|E|}{|V| - 1} = \frac{15}{9} = \frac{5}{3}.$$

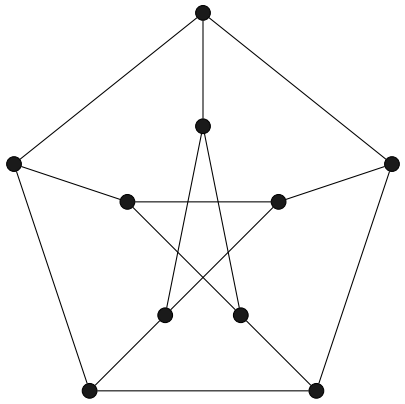


Figure 6.4: *Petersen graph*

One may check via direct computation of the density of subgraphs of the Peterson graph, that no subgraph is as dense as Peterson graph itself which proves that it must be irreducible.

Now, we assume $G = (V, E)$ is d-regular. Notice that $|E| = \frac{1}{2}d|V|$ in this case.

Claim 6.2.4. *Let G be a d-regular graph on $|V| = 2k$ vertices. If $2k-1$ and d are relatively prime, then G is irreducible.*

Proof. Indeed, since the condition in Proposition 6.2.2 can be translated in this case as follows - if dk and $(2k-1)$ are relatively prime, then G is irreducible. Since k and $2k-1$ are relatively prime for any k, then the statement about dk and $(2k-1)$ being relatively prime is true iff d and $(2k-1)$ are relatively prime. □

This gives rise to the following question.

Question 6.2.5. What is the likelihood of the event that $d(\leq 2k-1)$ and $2k-1$ are relatively prime for any positive integer k?

For simplicity, first we investigate the case when $d = 3$ and $k \geq 2$. The probability that $(2k-1)$ is divisible by 3 is equal $1/3$. This observation yields the following claim.

Claim 6.2.6. *Given a family of 3-regular graphs G on $2k$ vertices ($k \geq 2$), at least $2/3$ of them are irreducible.*

Now moving forward we will study two distinct cases - when d is even and when d is odd. If d is even, then we can represent it as

$$d = 2^{m_1} p_2^{m_2} ... p_s^{m_s},$$

where $p_i \neq 2$ for $i = 2, .., s$, are distinct prime numbers with m_i multiplicities correspondingly. In other words we represent d via its prime factorization.

Claim 6.2.7. *Let G be a d-regular graph on $2k$ vertices. Let d be an even number with the following prime factorization*

$$d = 2^{m_1} p_2^{m_2} ... p_s^{m_s}.$$

Then the likelihood that G is irreducible is bounded from below by the following number

$$1 - \left(\sum_{i=2}^{s} \frac{1}{p_i} - \sum_{i<j}^{s} \frac{1}{p_i p_j} - \sum_{i<j<k}^{s} \frac{1}{p_i p_j p_k} - ... - \frac{1}{p_2 ... p_s} \right).$$

Remark 6.2.8. Notice that in case when d is odd, then the likelihood that G is irreducible, is bounded below by

$$1 - \left(\sum_{i=1}^{s} \frac{1}{p_i} - \sum_{i<j}^{s} \frac{1}{p_i p_j} - \sum_{i<j<k}^{s} \frac{1}{p_i p_j p_k} - ... - \frac{1}{p_1 ... p_s} \right),$$

where we assume the following prime factorization for d

$$d = p_1^{m_1} p_2^{m_2} ... p_s^{m_s}.$$

Proof. The likelihood that $(2k - 1)$ is divisible by d is the same as saying it is divisible by either of the prime factors subtracted all the cases when j for $j = 2, ...s$ of them divide $(2k - 1)$. □

Bibliography

[1] N. Albin, J. Clemens, D. Hoare, P. Poggi-Corradini, B. Sit, and S. Tymochko. Fairest edge usage and minimum expected overlap for random spanning trees.

[2] N. Albin, J. Lind, A. Melikyan, and P. Poggi-Corradini. Fair spanning trees of maximal entropy. In progress.

[3] Nathan Albin and Pietro Poggi-Corradini. Minimal subfamilies and the probabilistic interpretation for modulus on graphs. *J. Anal.*, 24(2):183–208, 2016.

[4] Romas Aleliunas, Richard M. Karp, Richard J. Lipton, László Lovász, and Charles Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *20th Annual Symposium on Foundations of Computer Science (San Juan, Puerto Rico, 1979)*, pages 218–223. IEEE, New York, 1979.

[5] Bahman Bahmani, Ravi Kumar, and Sergei Vassilvitskii. Densest subgraph in streaming and mapreduce. *Proc. VLDB Endow.*, 5(5):454–465, January 2012.

[6] N. Biggs. *Algebraic Graph Theory*. Cambridge University Press, 1993.

[7] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[8] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[9] Andrei Broder and Eli Shamir. On the second eigenvalue of random regular graphs. In *Proceedings of the 28th Annual Symposium on Foundations of Computer Science*, SFCS '87, pages 286–294, Washington, DC, USA, 1987. IEEE Computer Society.

[10] Andrei Z. Broder and Anna R. Karlin. Bounds on the cover time. *J. Theoret. Probab.*, 2(1):101–120, 1989.

[11] Ashok K. Chandra, Prabhakar Raghavan, Walter L. Ruzzo, Roman Smolensky, and Prasoon Tiwari. The electrical resistance of a graph captures its commute and cover times. *computational complexity*, 6:312–340, 1996.

[12] Moses Charikar. Greedy approximation algorithms for finding dense components in a graph. In *Proceedings of the Third International Workshop on Approximation Algorithms for Combinatorial Optimization*, APPROX '00, pages 84–95, Berlin, Heidelberg, 2000. Springer-Verlag.

[13] Sunil Chopra. On the spanning tree polyhedron. *Oper. Res. Lett.*, 8(1):25–29, 1989.

[14] J. Clemens. Spanning tree modulus: deflation and a hierchial graph structure.

[15] Y. Peres D. Levin and E. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

[16] P. G. Doyle and Jessica L. Snell. Random walks and electric networks. 1987.

[17] Franois Fouss, Alain Pirotte, Jean-Michel Renders, and Marco Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 19, 2007.

[18] J. Friedman, J. Kahn, and E. Szemerédi. On the second eigenvalue of random regular graphs. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, pages 587–598, New York, NY, USA, 1989. ACM.

[19] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[20] Arpita Ghosh, Stephen Boyd, and Amin Saberi. Minimizing effective resistance of a graph. *SIAM Rev.*, 50(1):37–66, February 2008.

[21] Jean-Louis Goffin, Zhi-Quan Luo, and Yinyu Ye. Complexity analysis of an interior cutting plane method for convex feasibility problems. *SIAM J. on Optimization*, 6(3):638–652, March 1996.

[22] A. V. Goldberg. Finding a maximum density subgraph. Technical report, Berkeley, CA, USA, 1984.

[23] Robert M. Gray. *Entropy and Information Theory*. Springer Publishing Company, Incorporated, 2nd edition, 2011.

[24] Geoffrey Grimmett. *Probability on Graphs: Random Processes on Graphs and Lattices*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2 edition, 2018.

[25] A. Karr. Markov chains: Theory and applications (dean l. isaacson and richard w. madsen). *SIAM Review*, 20(3):606–607, 1978.

[26] G Kirchhoff. Ueber die auflosung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer strome gefuhrt wird. *Ann. Phys. und Chem.*, 72:497–508.

[27] Stephen J. Kirkland, Michael Neumann, and Bryan L. Shader. Distances in weighted trees and group inverse of laplacian matrices. 1997.

[28] K.C. Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming*, 90(1):1–25, 2001.

[29] Douglas Klein and Milan Randic. Resistance distance. *Journal of Mathematical Chemistry*, 12:81–95, 12 1993.

[30] I. V. Konnov. On convergence properties of a subgradient method. *Optimization Methods and Software*, 18(1):53–62, 2003.

[31] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*, volume 1st Ed. Cambridge University Press New York, 2017.

[32] Y.E. Nesterov. Minimization methods for nonsmooth convex and quasiconvex functions. *M*, 29:519–531.

[33] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 2005.

[34] M.E.J. Newman. The structure and function of complex networks. *Computer Physics Communications*, 147:40–45, 03 2003.

[35] Gyan Ranjan and Zhi-Li Zhang. Geometry of complex networks and topological centrality. *ArXiv*, abs/1107.0989, 2013.

[36] Gyan Ranjan, Zhi-Li Zhang, and Daniel L. Boley. Incremental computation of pseudo-inverse of laplacian. In *COCOA*, 2014.

[37] Jack Sherman and Winifred J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann. Math. Statist.*, 21(1):124–127, 03 1950.

[38] Nikhil Srivastava. Matrix tree theorems. 2009.

[39] Prasad Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4:101–109, 1991.

[40] David Bruce Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, STOC '96, pages 296–303, New York, NY, USA, 1996. ACM.

[41] Christopher S. Withers and Saralees Nadarajah. log det a = tr log a. *International Journal of Mathematical Education in Science and Technology*, 41(8):1121–1124, 2010.

[42] Wenjun Xiao and Ivan Gutman. Resistance distance and laplacian spectrum. *Theoretical Chemistry Accounts*, 110:284–289, 11 2003.

Appendix A

Convex Optimization

In this section we discuss convex optimization problems of the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0 \quad i = 1, ..., m, \\ & h_i(x) = 0 \quad i = 1, ..., p, \end{aligned}$$

where the functions $f_0, ..., f_m : \mathbb{R}^n \to \mathbb{R}$ are convex ,i.e.,

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

for all $x, y \in \mathbb{R}^n$ and for all $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$. We define the Lagrangian of this problem to be the following:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x).$$

We define the Lagrange dual function to be the following:

$$g(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \nu) = \inf_{x \in \mathbb{R}^n} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right).$$

Now we are ready to define the Lagrange dual problem:

$$\begin{aligned} &\text{maximize} \quad g(\lambda, \nu) \\ &\text{subject to} \quad \lambda \geq 0. \end{aligned}$$

In this context we refer to the original problem as the primal problem. Particularly, for the modulus problem we will have the following Lagrangian function:

$$\mathcal{L}(\rho, \lambda) := \sum_{e \in E} \sigma(e)|\rho(e)|^p + \sum_{\gamma \in \Gamma} \lambda(\gamma) \left(1 - \sum_{e \in E} \mathcal{N}(\gamma, e)\rho(e) \right).$$

In addition, we have the following dual function for the modulus problem:

$$g(\lambda) = \sum_{\gamma \in \Gamma} \lambda(\gamma) - (p-1) \sum_{e \in E} \sigma(e) \left(\frac{1}{p\sigma(e)} \sum_{\gamma \in \Gamma} \lambda(\gamma)\mathcal{N}(\gamma, e) \right)^{\frac{p}{p-1}}.$$

Below we discuss the KKT conditions for convex optimization problems. Let $\tilde{x}, \tilde{\lambda}$ and $\tilde{nu}$ be any points that satisfy the following KKT conditions:

$$\begin{aligned} \nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0 \\ f_i(\tilde{x}) \leq 0, \quad i = 1, ..., m \\ h_i(\tilde{x}) = 0, \quad i = 1, ..., p \\ \tilde{\lambda}_i \geq 0, \quad i = 1, ..., m \\ \tilde{\lambda}_i f_i(\tilde{x}) = 0, \quad i = 1, ..., m, \end{aligned}$$

then $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal.

Remark A.0.1. In order for a minimum $\tilde{x}$ to satisfy the KKT conditions, the convex problem should satisfy the following regularity conditions: there exists a point x such that $h(x) = 0$ and $f_i < 0$.

Appendix B

Quasiconvex Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *quasiconvex* if its domain and all its sublevel sets

$$S_\alpha = \{x \in \text{dom } f \mid f(x) \leq \alpha\},$$

for $\alpha \in \mathbb{R}$, are convex. A function f is quasiconcave if $-f$ is quasiconvex, i.e., every superlevel set $\{x \mid f(x) \geq \alpha\}$ is convex.

Example B.0.1. We consider the linear-fractional transformation

$$f(x) = \frac{a^T x + b}{c^T x + d}$$

with dom $f = \{x \mid c^T x + d > 0\}$. We notice that its α-sublevel set is convex since it is the intersection of an open halfspace and a closed halfspace.

$$\begin{aligned} S_\alpha &= \{x \mid c^T x + d > 0, (a^T x + b)/(c^T x + d) \leq \alpha\} \\ &= \{x \mid c^T x + d > 0, a^T x + b \leq \alpha(c^T x + d)\} \end{aligned}$$

However, the most commonly used characteristics for quasionvex functions is the property analogous to Jensen's inequality for convex functions. A function f is quasiconvex if and

only if dom f is convex and for any $x, y \in$ dom f and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\},$$

i.e., the value of the function on a segment does not exceed the maximum of its values at the endpoints.

Now, we introduce the first order conditions for quasiconvexity of a differentiable function. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then f is quasiconvex if and only if $textdomf$ is convex and for all $x, y \in$ dom f

$$f(y) \leq f(x) \Rightarrow \nabla f(x)^T (y - x) \leq 0.$$

This condition has the following geometric interpretation when $\nabla f(x) \neq 0$. It states that $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{y \mid f(y) \leq f(x)\}$, at the point x.

Remark B.0.2. From the optimization point of you the main difference between convex and quasiconvex functions is that $\nabla f(x) = 0$ condition does not imply that x is a global optimizer for f if f is quasiconvex unlike the case if it were convex.

Now suppose f is twice differentiable. If f is quasiconvex, then for all $x \in$ dom f, and all $y \in mathbbmR^n$, we have

$$y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y \geq 0.$$

Remark B.0.3. If we consider a quasiconvex function defined on $\mathbb{R}$, then this condition translates to the following

$$f'(x) = 0 \Rightarrow f''(x) \geq 0,$$

which can be interpreted as the second derivative being nonnegative at any point with zero slope.

Appendix C

Quasi-convexity of the energy

Claim C.0.1. *The energy can be computed by the following formula:*

$$\mathcal{E}_t = \frac{(1+t)^2 R_1^2}{(1+tR_1)^2} - \frac{2t}{1+tR_1} B + \frac{t^2}{(1+tR_1)^2} A + C.$$

Proof. We break the updated energy into two distinct pieces.

$$\mathcal{E}_t = \sigma_1^2 (1+t)^2 (u_{e_1}^T L_{\sigma_t}^+ u_{e_1})^2 + \sum_{e \neq e_1} \sigma(e)^2 (u_e^T L_{\sigma_t}^+ u_e)^2$$

Next we express $L_{\sigma_t}^+$ in terms of L_σ^+ as discussed above.

$$\begin{aligned}\mathcal{E}_t =& \sigma_1^2 (1+t)^2 \left(u_{e_1}^T \left(L_\sigma^+ - \frac{t\sigma_1}{1+t\sigma_1 R_1} L_\sigma^+ u_{e_1} u_{e_1}^T L_\sigma^+ \right) u_{e_1} \right)^2 \\ &+ \sum_{e \neq e_1} \sigma(e)^2 \left(u_e^T \left(L_\sigma^+ - \frac{t\sigma_1}{1+t\sigma_1 R_1} L_\sigma^+ u_{e_1} u_{e_1}^T L_\sigma^+ \right) u_e \right)^2\end{aligned}$$

Here we use the formula for the effective resistance on the edge e_1:

$$\begin{aligned}\mathcal{E}_t =&\sigma_1^2(1+t)^2\left(\frac{R_1}{1+t\sigma_1 R_1}\right)^2+\sum_{e\neq e_1}\sigma(e)^2\left(u_e^T L_\sigma^+ u_e-\frac{t\sigma_1}{1+t\sigma_1 R_1}(u_{e_1}^T L_\sigma^+ u_e)^2\right)^2\\ =&\sigma_1^2(1+t)^2\frac{R_1^2}{(1+t\sigma_1 R_1)^2}+\sum_{e\neq e_1}\sigma(e)^2\left(u_e^T L_\sigma^+ u_e\right)^2\\ &-\frac{2t\sigma_1}{1+t\sigma_1 R_1}\sum_{e\neq e_1}\sigma(e)^2(u_e^T L_\sigma^+ u_e)(u_{e_1}^T L_\sigma^+ u_e)^2+\frac{t^2\sigma_1^2}{(1+t\sigma_1 R_1)^2}\sum_{e\neq e_1}\sigma(e)^2(u_{e_1}^T L_\sigma^+ u_e)^4\end{aligned}$$

Finally, we use the notations introduced above to make further use of this formula more convenient.

$$\begin{aligned}\mathcal{E}_t =&\frac{\sigma_1^2(1+t)^2R_1^2}{(1+t\sigma_1 R_1)^2}-\frac{2t\sigma_1}{1+t\sigma_1 R_1}B+\frac{t^2\sigma_1^2}{(1+t\sigma_1 R_1)^2}A+C\\ =&\frac{(1+t)^2R_1^2}{(\sigma_1^{-1}+tR_1)^2}-\frac{2t}{\sigma_1^{-1}+tR_1}B+\frac{t^2}{(\sigma_1^{-1}+tR_1)^2}A+C\end{aligned}$$

Note that if we multiply all the weights by the same factor λ, as in $\sigma\mapsto\lambda\sigma$, then by Lemma 5.1.2, $\mathcal{E}_t$ remains constant. So we can assume that $\sigma_1=1$, and thus $R_1\leq 1$. This argument allows us to attain the desirable conclusion about the energy. □

Moving forward, we would like to compute the derivative of the energy, as our goal is to show the existence of the unique minimizer for the energy. The computations yield the following result.

$$\begin{aligned}\frac{d\mathcal{E}_t}{dt} =&2R_1^2\frac{(1+t)(1-R_1)}{(1+tR_1)^3}-\frac{2B}{(1+tR_1)^2}+\frac{2tA}{(1+tR_1)^3}\\ =&2\frac{(R_1^2(1-R_1)+A-R_1B)t+R_1^2(1-R_1)-B}{(1+tR_1)^3}\end{aligned}$$

Note that the numerator of $d\mathcal{E}_t/dt$ is a linear function and the denominator is always positive since $t\geq -1$ and $R_1\leq 1$. We claim that the slope of that linear function is positive, so that

$d\mathcal{E}_t/dt$ changes sign only once and goes from being negative to being positive.

Claim C.0.2. *The map $t \mapsto \mathcal{E}_t$ is quasiconvex.*

Proof. The slope of the numerator is $A + R_1(R_1(1 - R_1) - B)$. We will use the fact that (since $\sigma_1 = 1$)

$$\begin{aligned} R_1 =& \sum_{e \in E} \sigma(e)(u_e L_\sigma^+ u_{e_1}^T)^2 = (u_{e_1} L_\sigma^+ u_{e_1}^T)^2 + \sum_{e \neq e_1} \sigma(e)(u_{e_1} L_\sigma^+ u_{e_1}^T)^2 \\ =& R_1^2 + \sum_{e \neq e_1} \sigma(e)(u_{e_1} L_\sigma^+ u_{e_1}^T)^2 \end{aligned}$$

Hence by the definition of B,

$$\begin{aligned} R_1(1 - R_1) - B =& R_1 - R_1^2 - B \\ =& \sum_{e \neq e_1} \sigma(e)(1 - \sigma(e)R(e))(u_e L_\sigma^+ u_{e_1})^2 \geq 0, \end{aligned}$$

because $\sigma(e)R(e) \leq 1$ for every e. □

Moreover, the minimum is attained for

$$t^* = -\frac{R_1^2(1 - R_1) - B}{R_1^2(1 - R_1) + A - R_1 B}$$

Since $\mathcal{E}_t$ is only defined on $[-1, \infty)$, we want to check whether $t^* > 1$. This can be seen either by direct computation or by plugging $t = -1$ in and seeing that the numerator of $d\mathcal{E}_t/dt$ is negative there:

$$-R_1^2(1 - R_1) - A + R_1 B + R_1^2(1 - R_1) - B = -A - B(1 - R_1) \leq 0.$$

Now we fix two edges e_1 and e_2 and the corresponding vectors u_1 and u_2. Let $U = [u_1 u_2]$ and let $C = \mathrm{Diag}(\sigma_1 a, \sigma_2 b)$ for constants $a, b > 0$. Then the edge-conductances are perturbed

as follow, for $t \geq \min\{-1/a, -1/b\}$:

$$\sigma_t = \sigma + t(\sigma_1 a \mathbb{1}_{e_1} + \sigma_2 b \mathbb{1}_{e_2}),$$

and the Laplacian becomes,

$$L_{\sigma_t} = L_\sigma + t\sigma_1 a u_1 u_1^T + t\sigma_2 b u_2 u_2^T = L_\sigma + U(tC)U^T.$$

Here, first we use the representation of the pseudo-inverse of the Laplacian in terms of the inverse of rank-one updated Laplacian. Then we apply Woodbury identity for the inverse. Here is what we get as a result.

$$\begin{aligned}
(L+UCV)^+ &= \left(L + UCV + \frac{J}{N}\right)^{-1} - \frac{J}{N} \\
&= \left(L+\frac{J}{N}\right)^{-1} - \left(L+\frac{J}{N}\right)^{-1} U \Big(C^{-1} + V\Big(L+\frac{J}{N}\Big)^{-1} U\Big)^{-1} V\Big(L+\frac{L}{J}\Big)^{-1} - \frac{J}{N} \\
&= L^+ - \Big(L^+ + \frac{J}{N}\Big) U \Big(C^{-1} + V\Big(L^+ + \frac{J}{N}\Big)U\Big)^{-1} V\Big(L^+ + \frac{J}{N}\Big) \\
&= L^+ - \Big(L^+ + \frac{J}{N}\Big) U \Big(C^{-1} + VL^+U + V\frac{J}{N}U\Big)^{-1} V\Big(L^+ + \frac{J}{N}\Big) \\
&= L^+ - L^+U\Big(C^{-1} + VL^+U\Big)^{-1} V\Big(L^+ + \frac{J}{N}\Big) \\
&- \frac{J}{N} U\Big(C^{-1} + VL^+U\Big)^{-1} V\Big(L^+ + \frac{J}{N}\Big) \\
&= L^+ - L^+U\Big(C^{-1} + VL^+U\Big)^{-1} VL^+ - L^+U\Big(C^{-1} + VL^+U\Big)^{-1} VL\frac{J}{N} \\
&= L^+ - L^+U\Big(C^{-1} + VL^+U\Big)^{-1} VL^+
\end{aligned}$$

Therefore, we can use the following formula for $\text{Ł}^+_{\sigma t}$

$$L^+_{\sigma_t} = L^+_\sigma - L^+_\sigma U\left[(tC)^{-1} + U^T L_\sigma U\right]^{-1} U^T L^+_\sigma$$

As a result, we can write the energy as follows

$$\mathcal{E}(t) = \sum_{e \in E} \sigma_t(e)^2 (u_e^T L_{\sigma_t}^+ u_e)^2.$$

In order to simplify notation we write $A := \sigma_1 at$ and $B = \sigma_2 bt$. Note that

$$(tC)^{-1} + U^T L_\sigma U = \mathrm{Diag}(A^{-1}, B^{-1}) + \begin{bmatrix} R_1 & \nabla h_1(e_2) \\ \nabla h_1(e_2) & R_2 \end{bmatrix}.$$

So $\left[(tC)^{-1} + U^T L_\sigma U\right]^{-1}$ is equal to

$$\frac{1}{(A^{-1} + R_1)(B^{-1} + R_2) - |\nabla h_1|^2(e_2)} \begin{bmatrix} B^{-1} + R_2 & -\nabla h_1(e_2) \\ -\nabla h_1(e_2) & A^{-1} + R_1 \end{bmatrix}$$

which we write as $d^{-1}M$. Note that the denominator d is non-negative because $R_1 R_2 \geq |\nabla h_1|^2(e_2)$ by the Cauchy-Schwarz inequality.

The part of the energy corresponding to e_1 is $(\sigma_1 + A)^2$ times the square of

$$u_1^T \left(L_\sigma^+ - L_\sigma^+ U \left[(tC)^{-1} + U^T L_\sigma U\right]^{-1} U^T L_\sigma^+ \right) u_1 = R_1 - u_1^T L_\sigma^+ U d^{-1} M U^T L_\sigma^+ u_1$$

Note that $u_1^T L_\sigma^+ U = \begin{bmatrix} R_1 & \nabla h_1(e_2) \end{bmatrix}$. So the second term is d^{-1} times

$$R_1^2 B^{-1} + |\nabla h_1|^2(e_2) A^{-1} + R_1(R_1 R_2 - |\nabla h_1|^2(e_2))$$

After finding common denominators, we get d^{-1} times

$$R_1 A^{-1} B^{-1} + (R_1 R_2 - |\nabla h_1|^2(e_2)) A^{-1}$$

So setting $\delta = R_1 R_2 - |\nabla h_1|^2(e_2)$, this part of the energy is

$$(\sigma_1 + A)^2 \left(\frac{R_1 + \delta B}{1 + R_1 A + R_2 B + \delta AB} \right)^2 \left(\frac{\sigma_1 R_1 + R_1 A + \sigma_1 \delta B + \delta AB}{1 + R_1 A + R_2 B + \delta AB} \right)^2$$

which again tends to 1 as $t \to \infty$.

Appendix D

One edge perturbation

Here we provide several different approaches to perturbing one edge, with the intention to generalize one of them to the case of perturbing several edges simultaneously, which would allow to take oblique directions.

D.0.1 Probabilistic approach

In this section we discuss a probabilistic approach to describe the change in energy while updating the weight on a single edge of the graph. We enumerate the edges $e_1, \dots$. Write σ_1 for the weight of e_1. For $e \neq e_1$, write $\sigma(e)$. We increase the weight on edge e_1 from σ_1 to $\sigma_1 + k$.

Remark D.0.1. If σ_1 and k are integer, we can think of e_1 as having σ_1 multi-edges and adding k new ones. However, when computing edge probabilities we still think of simple graphs. Namely, adding multi-edges increases the number of trees that use the simple edge e_1, where all the multi-edges are collapsed.

We write σ for the old weights and σ' for the new weights. Then

$$\begin{aligned} p := \mathbb{P}(e_1 \in T \mid \sigma) &= \frac{\sum_{T \ni e_1} \prod_{e \in T} \sigma(e)}{\sum_T \prod_{e \in T} \sigma(e)} \\ &= \frac{\sigma_1 \sum_{T \ni e_1} \prod_{e_1 \neq e \in T} \sigma(e)}{\sigma_1 \sum_{T \ni e_1} \prod_{e_1 \neq e \in T} \sigma(e) + \sum_{T \not\ni e_1} \prod_{e \in T} \sigma(e)} \\ &= \frac{\sigma_1 A}{\sigma_1 A + B} \end{aligned}$$

As a result we can express B in terms of A, σ_1 and p.

$$B = \sigma_1 A(p^{-1} - 1).$$

On the other hand, if we consider the edge-usage probabilities for the updated edge after the update, we will have the following

$$\begin{aligned} \mathbb{P}(e_1 \in T \mid \sigma') &= \frac{\sum_{T \ni e_1} \prod_{e \in T} \sigma'(e)}{\sum_T \prod_{e \in T} \sigma'(e)} \\ &= \frac{(\sigma_1 + k) \sum_{T \ni e_1} \prod_{e_1 \neq e \in T} \sigma(e)}{(\sigma_1 + k) \sum_{T \ni e_1} \prod_{e_1 \neq e \in T} \sigma(e) + \sum_{T \not\ni e_1} \prod_{e \in T} \sigma(e)} \\ &= \frac{(\sigma_1 + k)A}{(\sigma_1 + k)A + B} \\ &= \frac{\sigma_1 + k}{\sigma_1 p^{-1} + k} \end{aligned}$$

Assuming $\sigma_1 = 1$ and writing $\mathcal{E}_1(k) = \mathbb{P}(e_1 \in T \mid \sigma')^2$, we get

$$\mathcal{E}_1(k) = \left(p \frac{1+k}{1+pk}\right)^2 = \left(1 - \frac{1-p}{1+pk}\right)^2.$$

Note $\mathcal{E}_1(-1) = 0$ and $\lim_{k \to \infty} \mathcal{E}_1(k) = 1$.

Now, we discuss the case when $e \neq e_1$.

$$p(e) := \mathbb{P}(e \in T \mid \sigma) = \frac{\sum_{T \ni e} \prod_{e' \in T} \sigma(e')}{\sum_{T} \prod_{e' \in T} \sigma(e')}$$

The numerator of this expression can be simplified to

$$\sigma(e)\sigma_1 \sum_{e_1 \in T \ni e} \prod_{e' \neq e, e_1 \in T} \sigma(e') + \sigma(e) \sum_{e_1 \notin T \ni e} \prod_{e' \neq e \in T} \sigma(e')$$

and the denominator is of the following form

$$\begin{aligned}\sigma(e)\sigma_1 \sum_{e_1 \in T \ni e} \prod_{e' \neq e, e_1 \in T} \sigma(e') + \sigma(e) \sum_{e_1 \notin T \ni e} \prod_{e' \neq e \in T} \sigma(e') \\ + \sigma_1 \sum_{e_1 \in T \not\ni e} \prod_{e_1 \neq e' \in T} \sigma(e') + \sum_{e_1 \notin T \not\ni e} \prod_{e' \in T} \sigma(e')\end{aligned}$$

Eventually, we combine them to obtain the following ratio

$$p(e) = \frac{A(e_1, e) + A(\overline{e_1}, e)}{A(e_1, e) + A(\overline{e_1}, e) + A(e_1, \overline{e}) + A(\overline{e_1}, \overline{e})} \tag{D.0.1}$$

where the notation $A(e, e', \ldots, \overline{f}, \overline{f'}, \ldots)$ represents the weight of all the trees that include edges $e, e', \ldots$ but not $f, f' \ldots$.

To simplify further, let $a := A(e_1, e)$, $b := A(e_1, \overline{e})$, $c := A(\overline{e_1}, e)$, and $d := A(\overline{e_1}, \overline{e})$. Then (D.0.1) can be written as

$$a + b + c + d = p(e)^{-1}(a + c). \tag{D.0.2}$$

Also, when passing from σ to σ', a changes to $(1 + k/\sigma_1)a$ and b changes to $(1 + k/\sigma_1)b$.

Therefore,

$$\begin{aligned}\mathbb{P}(e \in T \mid \sigma') &= \frac{(1+k/\sigma_1)A(e_1,e) + A(\overline{e_1},e)}{(1+k/\sigma_1)A(e_1,e) + A(\overline{e_1},e) + (1+k/\sigma_1)A(e_1,\overline{e}) + A(\overline{e_1},\overline{e})} \\ &= \frac{ka/\sigma_1 + a + c}{ka/\sigma_1 + kb/\sigma_1 + a + b + c + d}\end{aligned}$$

Assuming without loss of generality that $\sigma_1 = 1$ and using (D.0.2)

$$\begin{aligned}\mathbb{P}(e \in T \mid \sigma') &= \frac{ka + a + c}{k(a+b) + p(e)^{-1}(a+c)} \\ &= \frac{a}{a+b}\left[1 + \frac{\frac{a+b}{a}(a+c) - p(e)^{-1}(a+c)}{k(a+b) + p(e)^{-1}(a+c)}\right] \\ &= \frac{a}{a+b} + \frac{p(e) - a(a+b)^{-1}}{k(a+b) + p(e)^{-1}(a+c)}p(e)^{-1}(a+c)\end{aligned}$$

We claim that $p(e) - a(a+b)^{-1} \geq 0$, so that, as $k \to \infty$, $\mathbb{P}(e \in T \mid \sigma')$ is convex and decreases, with horizontal asymptote $a/(a+b)$. To see this recall that $p(e) = \mathbb{P}_\sigma(e \in T)$ and note that

$$\frac{a}{a+b} = \frac{A(e_1,e)}{A(e_1,e) + A(e_1,\overline{e})} = \mathbb{P}_\sigma\left(e \in T \mid e_1 \in T\right).$$

However,

$$\mathbb{P}_\sigma(e \in T) \geq \mathbb{P}_\sigma\left(e \in T \mid e_1 \in T\right)$$

because random trees are negative correlated, see (cite grimmett) Intuitively, if one already knows that $e_1 \in T$, then it's less likely that $e \in T$.

D.0.2 Effective resistance version

In this section we will be working with the updated effective resistances on the edges of the graph as we update the weight on one of the edges while keeping the weights on the remaining edges the same. Enumerate the edges $e_1, \ldots$. Write R_1 and σ_1 for the effective resistance and the conductance of e_1. For $e \neq e_1$, write $R(e)$ and $\sigma(e)$. Recall that R_1 equals the voltage

drop across e_1 necessary to pass a unit of current across the network. Let $h_1 : V \to \mathbb{R}$ be this voltage potential and for every edge $e = \{x, y\}$, write $|\nabla h_1|(e) = |h(x) - h(y)|$. Then

$$R_1 = |\nabla h_1|(e_1). \tag{D.0.3}$$

Moreover, R_1 is also the dissipated power in the system,

$$R_1 = \sum_{e \in E} \sigma(e) |\nabla h_1|^2(e). \tag{D.0.4}$$

One way to see this is to note that h_1/R_1 is a unit potential drop, so its energy is effective conductance, i.e.,

$$\frac{1}{R_1} = \sum_{e \in E} \sigma(e) \frac{|\nabla h_1|^2(e)}{R_1^2},$$

which can be seen to be equivalent to (D.0.4).

We now let $t \in [-1, \infty)$ and write $\sigma_t = \sigma + t\sigma_1 \mathbb{1}_{e_1}$. This gives rise to the energy as a function of t:

$$\mathcal{E}(t) = \sum_{e \in E} \sigma_t(e)^2 R_t(e)^2.$$

where R_t is effective resistance with respect to σ_t. To understand this sum we need to compute $R_t(e)$.

For e_1, think of edge e_1 as being in parallel with the rest of the network. Let $\tilde{R}$ be the effective resistance of the rest of the network. Then, by the parallel rule

$$R_t(e_1)^{-1} = \tilde{R}^{-1} + \sigma_1(1 + t).$$

and when $t = 0$, we have $R_1^{-1} = \tilde{R}^{-1} + \sigma_1$, so

$$R_t(e_1)^{-1} = R_1^{-1} + t\sigma_1,$$

i.e.

$$R_t(e_1) = \frac{1}{R_1^{-1} + t\sigma_1} \tag{D.0.5}$$

Now assume that $e \neq e_1$.

Fix an orientation on the edges. If $e = (x, y)$, write $u_e = \delta_y - \delta_x$ Recall the Laplacian is:

$$L_\sigma = \sum_{e \in E} \sigma(e) u_e u_e^T$$

so that

$$L_{\sigma_t} = L_\sigma + t\sigma_1 u_{e_1} u_{e_1}^T.$$

By Shermann-Morrison:

$$L_{\sigma_t}^+ = L_\sigma^+ - \frac{1}{(t\sigma_1)^{-1} + R_1} L_\sigma^+ u_{e_1} u_{e_1}^T L_\sigma^+$$

Therefore, for $e \neq e_1$ we have

$$\begin{aligned}
R_t(e) &= u_e^T L_{\sigma_t}^+ u_e = u_e^T L_\sigma^+ u_e - \frac{1}{(t\sigma_1)^{-1} + R_1} u_e^T L_\sigma^+ u_{e_1} u_{e_1}^T L_\sigma^+ u_e \\
&= R(e) - \frac{1}{(t\sigma_1)^{-1} + R_1} (u_e^T L_\sigma^+ u_{e_1})^2. \\
&= R(e) - \frac{1}{(t\sigma_1)^{-1} + R_1} |\nabla h_1|^2(e). \\
\mathcal{E}(t) &= (1+t)^2 \sigma_1^2 R_t(e_1)^2 + \sum_{e \neq e_1} \sigma(e)^2 R_t(e)^2 \\
&= \frac{(1+t)^2 \sigma_1^2}{(R_1^{-1} + t\sigma_1)^2} + \sum_{e \neq e_1} \sigma(e)^2 \left(R(e) - \frac{1}{(t\sigma_1)^{-1} + R_1} |\nabla h_1|^2(e) \right)^2 \\
&= \frac{(1+t)^2 \sigma_1^2}{(R_1^{-1} + t\sigma_1)^2} + \sum_{e \neq e_1} \sigma(e)^2 \left(R(e) - \frac{t\sigma_1 R_1^{-1}}{R_1^{-1} + t\sigma_1} |\nabla h_1|^2(e) \right)^2
\end{aligned}$$

We make the change variables $s = R_1^{-1} + t\sigma_1$. Note that $t = \sigma_1^{-1} \left[s - R_1^{-1} \right]$. In particular, $t \geq -1$ if and only if $s \geq R_1^{-1}(1 - \sigma_1 R_1) \geq 0$. So if $\phi(s) := \mathcal{E}(t(s))$, then $\phi'(s) = \mathcal{E}'(t(s)) \frac{dt}{ds}$

and

$$\frac{dt}{ds} = \sigma_1^{-1} \geq 0.$$

Since we are only interested in the sign of $\mathcal{E}'$, we can without loss of generality, study the sign of ϕ' instead. Note that

$$\begin{aligned}
\phi(s) =& s^{-2}(\sigma_1 - R_1^{-1} + s)^2 + \sum_{e \neq e_1} \sigma(e)^2 \left(R(e) - s^{-1}(s - R_1^{-1})R_1^{-1}|\nabla h_1|^2(e)\right)^2 \\
=& R_1^{-2}(R_1 - s^{-1}(1 - \sigma_1 R_1))^2 + \sum_{e \neq e_1} \sigma(e)^2 \left(R(e) - (1 - s^{-1}R_1^{-1})R_1^{-1}|\nabla h_1|^2(e)\right)^2
\end{aligned}$$

Here we take the derivative in s and simplify it to a linear function in s with slope:

$$\begin{aligned}
\frac{d\phi(s)}{ds} =& 2R_1^{-2}(R_1 - (1 - \sigma_1 R_1)s^{-1})(1 - \sigma_1 R_1)s^{-2} \\
& - 2\sum_{e \neq e_1} \sigma(e)^2 \left(R(e) - (1 - s^{-1}R_1^{-1})R_1^{-1}|\nabla h_1|^2(e)\right) R_1^{-2}|\nabla h_1|^2(e)s^{-2} \\
\frac{s^3}{2}\frac{d\phi(s)}{ds} =& R_1^{-2}(sR_1 - (1 - \sigma_1 R_1))(1 - \sigma_1 R_1) \\
& - \sum_{e \neq e_1} \sigma(e)^2 \left(sR(e) - (s - R_1^{-1})R_1^{-1}|\nabla h_1|^2(e)\right) R_1^{-2}|\nabla h_1|^2(e) \\
m :=& R_1^{-1}(1 - \sigma_1 R_1) - R_1^{-2}\sum_{e \neq e_1} \sigma(e)^2 R(e)|\nabla h_1|^2(e) \\
& + R_1^{-3}\sum_{e \neq e_1} \sigma(e)^2 |\nabla h_1|^4(e) \\
\geq& R_1^{-1}(1 - \sigma_1 R_1) - R_1^{-2}\sum_{e \neq e_1} \sigma(e)|\nabla h_1|^2(e) && \text{(since } \sigma(e)R(e) \leq 1 \text{)} \\
=& R_1^{-1}(1 - \sigma_1 R_1) - R_1^{-2}(R_1 - \sigma_1 R_1^2) \\
=& R_1^{-2}[R_1(1 - \sigma_1 R_1) - (R_1 - \sigma_1 R_1^2)] = 0.
\end{aligned}$$

D.0.3 Calculus version

In this section, we apply calculus version analysis of the energy as we increase the weight on one edge of the graph only. As before, we have

$$\begin{aligned}\mathcal{E}(t) &= \sum_{e\in E} \left(\sigma_t(e)R_t(e)\right)^2 \\ &= \left(\sigma_t(e_1)R_t(e_1)\right)^2 + \sum_{e\neq e_1} \left(\sigma(e)R_t(e)\right)^2 = f(t) + g(t)\end{aligned}$$

Let $f(t)$ be the part of the energy contributed by e_1 and let $g(t)$ be the remaining part. The conductance on edge e_1 increases linearly $\sigma_t(e_1) = \sigma_1(1+t)$, for $t \geq -1$. This causes every effective resistance $R_t(E)$ to decrease. Indeed,

$$0 \leq R_t(e) = R(e) - \frac{|\nabla h_1(e)|^2}{\frac{1}{\sigma_1 t} + R_1} \leq R(e) \tag{D.0.6}$$

and

$$R_t'(e) = -\frac{\sigma_1 |\nabla h_1(e)|^2}{(1+\sigma_1 R_1 t)^2} \leq 0 \quad \text{and} \quad \lim_{t\to\infty} R_t'(e) = 0. \tag{D.0.7}$$

Note that the pole is at $t_0 = -(\sigma_1 R_1)^{-1}$ which is less than -1 since $\sigma_1 R_1 \leq 1$. Also,

$$R_{-1}(e) = R(e) + \frac{\sigma_1 |\nabla h_1(e)|^2}{1-\sigma_1 R_1}$$

and

$$\lim_{t\to\infty} R_t(e) = R(e) - \frac{|\nabla h_1(e)|^2}{R_1} \geq 0.$$

where the latter is non-negative because $R(e)R_1 \geq |\nabla h_1(e)|^2$, by the Cauchy-Schwarz inequality.

Also we have,

$$\lim_{t\to\infty} R_t'(e) = 0 \quad \text{and} \quad R_{-1}'(e) = -\frac{\sigma_1 |\nabla h_1(e)|^2}{(1-\sigma_1 R_1)^2} \tag{D.0.8}$$

Finally,

$$R_t''(e) = \frac{2\sigma_1 R_1 |\nabla h_1(e)|^2}{(1+\sigma_1 R_1 t)^3} \geq 0. \tag{D.0.9}$$

So for every edge, $R_t(e)$ is a convex decreasing function with a horizontal asymptote at $+\infty$.

When $e = e_1$, from (D.0.6), we have

$$R_t(e_1) = R_1 - \frac{R_1^2}{\frac{1}{\sigma_1 t} + R_1} = \frac{R_1}{1 + \sigma_1 R_1 t} \tag{D.0.10}$$

In particular,

$$R_{-1}(e_1) = \frac{R_1}{1 - \sigma_1 R_1} \quad \text{and} \quad \lim_{t\to\infty} R_t(e_1) = R_1 - \frac{R_1^2}{R_1} = 0,$$

which makes sense because the edge-conductance of e_1 tends to infinity.

In fact,

$$\lim_{t\to\infty} \sigma_t(e_1) R_t(e_1) = 1. \tag{D.0.11}$$

Also,

$$R_t'(e_1) = -\frac{\sigma_1 R_1^2}{(1 + \sigma_1 R_1 t)^2} = -\sigma_1 R_t(e_1)^2 \tag{D.0.12}$$

As a result we see below that the derivative of the per-edge effective resistance is non-negative.

$$(\sigma_t(e_1) R_t(e_1))' = \sigma_1 R_t(e_1) \left(1 - \sigma_t(e_1) R_t(e_1)\right) \geq 0, \tag{D.0.13}$$

meaning that $\sigma_t(e_1) R_t(e_1)$ monotonically increases to 1. Finally,

$$R_t''(e_1) = \frac{2\sigma_1^2 R_1^3}{(1 + \sigma_1 R_1 t)^3} = 2\sigma_1^2 R_t(e_1)^3 \tag{D.0.14}$$

We want to show that

$$\mathcal{E}'(t) = f'(t) + g'(t)$$

has exactly one zero in $[-1, \infty)$ and goes from being negative to being positive.

Note that

$$g'(t) = 2 \sum_{e \neq e_1} \sigma(e)^2 R_t(e) R_t'(e) \leq 0,$$

and by (D.0.9),

$$g''(t) = 2\sum_{e\neq e_1} \sigma(e)^2 \left(R_t'(e)^2 + R_t(e)R_t''(e)\right) \geq 0.$$

So $g'(t)$ is negative and increasing, and by (D.0.7) has a horizontal asymptote at $y = 0$.

On the other hand, by (D.0.7) and (D.0.12),

$$\begin{aligned} f'(t) &= 2\sigma_t(e_1)R_t(e_1)\left(\sigma_t'(e_1)R_t(e_1) + \sigma_t(e_1)R_t'(e_1)\right) \\ &= 2\sigma_t(e_1)R_t(e_1)\left(\sigma_1 R_t(e_1) - \sigma_1\sigma_t(e_1)R_t(e_1)^2\right) \\ &= 2\sigma_1\sigma_t(e_1)R_t(e_1)^2\left(1 - \sigma_t(e_1)R_t(e_1)\right) \geq 0. \end{aligned}$$

So by (D.0.11) and (D.0.10),

$$\lim_{t\to\infty} f'(t) = 0. \tag{D.0.15}$$

Also,

$$\begin{aligned} f''(t) &= 2\left(\sigma_1 R_t(e_1) + \sigma_t(e_1)R_t'(e_1)\right)^2 + 2\sigma_t(e_1)R_t(e_1)\left(2\sigma_1 R_t'(e_1) + \sigma_t(e_1)R_t''(e_1)\right) \\ &= 2\sigma_1^2 R_t(e_1)^2\left(1 - \sigma_t(e_1)R_t(e_1)\right)\left(1 - 3\sigma_t(e_1)R_t(e_1)\right) \end{aligned}$$

Note that $f''(-1) > 0$. Also, by (D.0.11) and (D.0.13), $\sigma_t(e_1)R_t(e_1)$ monotonically increases to 1. So, as $t \to \infty$, $f''(t)$ goes from positive to negative. Since g'' is always positive, we have that either $\mathcal{E}''$ is always positive, or it changes sign only once and goes from positive to negative. Moreover, by (D.0.7) and (D.0.15), $\mathcal{E}'(t)$ has horizontal asymptote $y = 0$ as well. Moreover, since $g'(-1) \leq 0$ and $f'(-1) = 0$, we have $\mathcal{E}'(-1) \leq 0$.

This means that $\mathcal{E}$ does have an inflection point, and the derivative there must be positive. So we find that $\mathcal{E}$ has a unique minimum and the function is quasiconvex.