RANDOM WALKS ON THRESHOLD GRAPHS

Project 4 - FUSRP 2023

FIELDS

August 9, 2023

Alexander Low Fung[1], Amy Mann[2], Andrei Parfeni[3], Giovanni Tedesco[4]

[1]San Francisco State University

[2]University of Toronto

[3]Yale University

[4]University of Toronto Mississauga

Background

Motivation

The problem

You are an exceptionally bad driver. You want to go to the Fields Institute, but you don't know how to get there, and you don't know how to use Google Maps.

How can you design the Toronto road system so you can get to your desired destination the fastest?

How can you design the Toronto road system so you can get to your desired destination the fastest?

The idea

- Model the situation using a Markov chain.

How can you design the Toronto road system so you can get to your desired destination the fastest?

The idea

- Model the situation using a Markov chain.
- Given the graph G whose vertices represent the states and whose edges represent the possible changes, we define the Markov chain such that no matter what vertex you are at, you have an equal probability of transitioning to any of the vertex's neighbors.

How can you design the Toronto road system so you can get to your desired destination the fastest?

The idea

- Model the situation using a Markov chain.
- Given the graph G whose vertices represent the states and whose edges represent the possible changes, we define the Markov chain such that no matter what vertex you are at, you have an equal probability of transitioning to any of the vertex's neighbors.
- Look at the short-term behavior of the Markov chain to figure out the expected time elapsed before you first reach the state which represents the Fields Institute.

Kemeny's Constant

Mean First Passage Time

The mean first passage time from i to j is the expected number of time-steps elapsed before the system reaches state j, given that it begins in state i. It is denoted by $m_{i,j}$.

Kemeny's Constant

Mean First Passage Time

The mean first passage time from i to j is the expected number of time-steps elapsed before the system reaches state j, given that it begins in state i. It is denoted by $m_{i,j}$.

Preliminary Definition of Kemeny's Constant

Let T be the transition matrix for a Markov chain with n states, with stationary vector w and mean first passage times $m_{i,j}$. Define $\kappa_i = \sum_{j=1, j\neq i}^{n} w_j m_{i,j}$.

Kemeny's Constant

Mean First Passage Time

The mean first passage time from i to j is the expected number of time-steps elapsed before the system reaches state j, given that it begins in state i. It is denoted by $m_{i,j}$.

Preliminary Definition of Kemeny's Constant

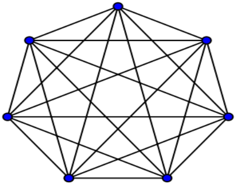
Let T be the transition matrix for a Markov chain with n states, with stationary vector w and mean first passage times $m_{i,j}$. Define $\kappa_i = \sum_{j=1, j\neq i}^{n} w_j m_{i,j}$.

This is a constant (it does not depend on the choice of i)!

Extremal values of Kemeny's constant

Smallest and largest Kemeny's constants

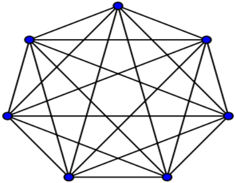
- The smallest Kemeny's constant on n vertices is $\mathcal{K}(K_n) = n - 2 + \frac{1}{n}$, corresponding to the complete graph.

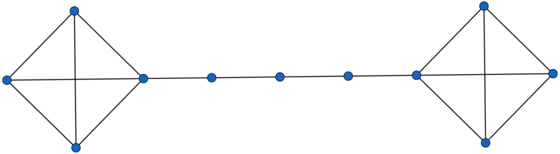


Extremal values of Kemeny's constant

Smallest and largest Kemeny's constants

- The smallest Kemeny's constant on n vertices is $\mathcal{K}(K_n) = n - 2 + \frac{1}{n}$, corresponding to the complete graph.
- (conjecture) The maximal Kemeny's constant on n vertices is $\mathcal{K}(G) = \frac{1}{54}n^3 + \mathcal{O}(n^2)$, corresponding to a "bridge" between two complete subgraphs.





A Graph Theory Interlude

Let G be an undirected connected simple graph on n vertices. Let $V(G) = \{v_1, \ldots, v_n\}$, with corresponding degrees $d_1, \cdots, d_n$.

Degree Matrix

$$D := \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{pmatrix}$$

In other words, $D = diag(d)$

A Graph Theory Interlude

Let G be an undirected connected simple graph on n vertices. Let $V(G) = \{v_1, \ldots, v_n\}$, with corresponding degrees $d_1, \cdots, d_n$.

Adjacency Matrix

We define the adjacency matrix A entry-wise as

$$A_{i,j} := \begin{cases} 1 & \text{if there is an edge between } v_i \text{ and } v_j \\ 0 & \text{otherwise} \end{cases}$$

In particular, this matrix illustrates what vertices are connected to each other. Notice also that $d = A \cdot \mathbf{1}$

Laplacian matrix

We define the Laplacian matrix L as $L = D - A$.

Computing Kemeny's constant

Second Definition of Kemeny's Constant

Suppose that G is a connected graph on n vertices and let $T = D^{-1}A$ denote its transition matrix. Suppose that $1, \lambda_2 \ldots, \lambda_n$ are the eigenvalues of T. Then we can (alternatively) define Kemeny's constant of G as

$$\mathcal{K}(G) = \sum_{j=2}^{n} \frac{1}{1-\lambda_j}$$

Computing Kemeny's constant

Second Definition of Kemeny's Constant

Suppose that G is a connected graph on n vertices and let $T = D^{-1}A$ denote its transition matrix. Suppose that $1, \lambda_2 \dots, \lambda_n$ are the eigenvalues of T. Then we can (alternatively) define Kemeny's constant of G as

$$\mathcal{K}(G) = \sum_{j=2}^{n} \frac{1}{1-\lambda_j}$$

The problem (part 2)

- Kemeny's constant is much simpler to calculate for graphs that are relatively regular (the degrees of the vertices are similar).

Computing Kemeny's constant

Second Definition of Kemeny's Constant

Suppose that G is a connected graph on n vertices and let $T = D^{-1}A$ denote its transition matrix. Suppose that $1, \lambda_2 \ldots, \lambda_n$ are the eigenvalues of T. Then we can (alternatively) define Kemeny's constant of G as

$$\mathcal{K}(G) = \sum_{j=2}^{n} \frac{1}{1-\lambda_j}$$

The problem (part 2)

- Kemeny's constant is much simpler to calculate for graphs that are relatively regular (the degrees of the vertices are similar).
- But what happens when they are not?

Threshold Graphs

Definition

A *threshold graph* is a graph that can be constructed from one vertex by repeatedly adding an isolated vertex or a dominating vertex. We can represent a threshold graph by a binary sequence $(b_1, \ldots, b_n)$ as follows:

- Start with a vertex v_0.
- For i from 1 to n, if $b_i = 0$ then we add an isolated vertex v_i to the graph. If $b_i = 1$ then we add a dominating vertex v_i, that is, we add edges joining v_i to each of $v_0, \ldots, v_{i-1}$.

We require that $b_n = 1$, so that the threshold graph is connected.

Threshold Graph Example

Threshold graph corresponding to the code 0011101

For example, the construction sequence $(0, 0, 1, 1, 1, 0, 1)$ represents the following threshold graph.

Threshold Graph Example

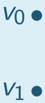
Threshold graph corresponding to the code 0011101

For example, the construction sequence $(0, 0, 1, 1, 1, 0, 1)$ represents the following threshold graph.

Threshold Graph Example

Threshold graph corresponding to the code 0011101

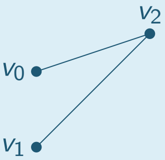
For example, the construction sequence $(0, 0, 1, 1, 1, 0, 1)$ represents the following threshold graph.



Threshold Graph Example

Threshold graph corresponding to the code 0011101

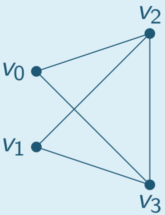
For example, the construction sequence $(0, 0, 1, 1, 1, 0, 1)$ represents the following threshold graph.



Threshold Graph Example

Threshold graph corresponding to the code 0011101

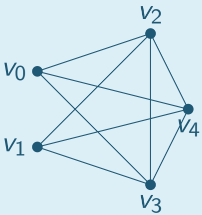
For example, the construction sequence $(0, 0, 1, 1, 1, 0, 1)$ represents the following threshold graph.



Threshold Graph Example

Threshold graph corresponding to the code 0011101

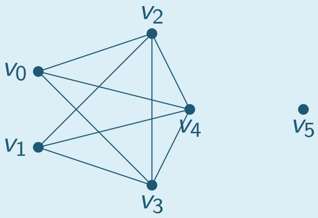
For example, the construction sequence $(0, 0, 1, 1, 1, 0, 1)$ represents the following threshold graph.



Threshold Graph Example

Threshold graph corresponding to the code 0011101

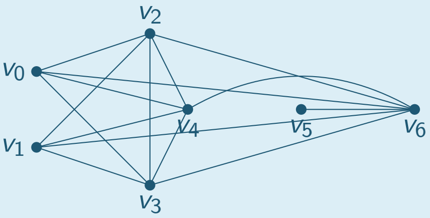
For example, the construction sequence $(0, 0, 1, 1, 1, 0, 1)$ represents the following threshold graph.



Threshold Graph Example

Threshold graph corresponding to the code 0011101

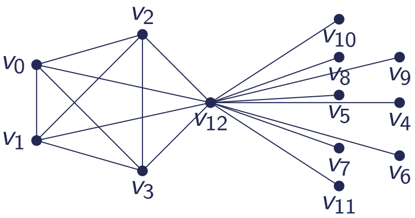
For example, the construction sequence $(0,0,1,1,1,0,1)$ represents the following threshold graph.



Results

Main Question

What threshold graph on n vertices maximizes Kemeny's constant?



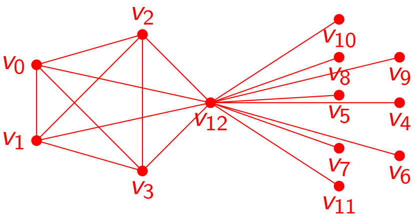
A Pineapple!

Conjecture

The Threshold graph G generated by the code

$$C = 0\underbrace{1\dots1}_{r \text{ times}}\ \underbrace{0\dots0}_{n-r-2 \text{ times}}\ 1$$

where $r \in \{\lfloor\sqrt{2n}\rfloor - 2, \lfloor\sqrt{2n}\rfloor - 1\}$, maximises Kemeny's constant for a code of length n. These are called **Pineapple Graphs**!



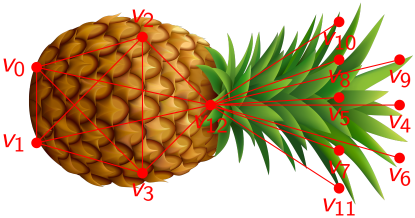
A Pineapple!

Conjecture

The Threshold graph G generated by the code

$$C = 0\underbrace{1\ldots1}_{r \text{ times}}\ \underbrace{0\ldots0}_{n-r-2 \text{ times}}\ 1$$

where $r \in \{\lfloor\sqrt{2n}\rfloor - 2, \lfloor\sqrt{2n}\rfloor - 1\}$, maximises Kemeny's constant for a code of length n. These are called **Pineapple Graphs**!



A Pineapple!

Conjecture

The Threshold graph G generated by the code

$$C = 0\underbrace{1\ldots1}_{r\text{ times}}\ \underbrace{0\ldots0}_{n-r-2\text{ times}}\ 1$$

where $r \in \{\lfloor\sqrt{2n}\rfloor - 2, \lfloor\sqrt{2n}\rfloor - 1\}$, maximises Kemeny's constant for a code of length n. These are called **Pineapple Graphs**!

Conjectured maximal value

From this code, we would obtain:

$$\max_{|V(G)|=n} \mathcal{K}(G) = n + \frac{\sqrt{n}}{2} + \mathcal{O}(1)$$

Our Approaches

Two main approaches:

1. Linear Algebra Approach
2. Combinatorial Approach

Linear Algebra Approach

Commutativity of the Laplacians

Theorem

Suppose that G_1 and G_2 are threshold graphs on n vertices, and let L_1, L_2 denote their respective Laplacians. Then we have that

$$L_1 L_2 = L_2 L_1$$

Commutativity of the Laplacians

Theorem

Suppose that G_1 and G_2 are threshold graphs on n vertices, and let L_1, L_2 denote their respective Laplacians. Then we have that

$$L_1 L_2 = L_2 L_1$$

Corollary

There exists a unitary real orthogonal matrix U which simultaneously diagonalizes the Laplacians of all Threshold graphs on n vertices. Its columns are common eigenvectors of all these Laplacians.

The U matrix

Structure and entries of U

The universal diagonalizing matrix U has a nice and useful structure for computations of Kemeny's constant:

$$e_i^T U e_j = \begin{cases} 0 & \text{if } i+j \geq n+3 \\ \frac{1}{\sqrt{n}} & \text{if } j = 1 \\ \frac{1}{\sqrt{(n+1-j)(n+2-j)}} & \text{if } i+j \leq n+1 \\ -\sqrt{\frac{n+1-j}{n+2-j}} & \text{if } i+j = n+2 \end{cases}$$

Resulting expressions for Kemeny's Constant

Expression 1

Let G be a threshold graph with n vertices, m be the number of its edges, d be its degree vector (ordered by the code), λ be the vector of eigenvalues of its Laplacian matrix (also ordered by the code), c be the code vector, and let U be the unitary diagonalizing matrix. Then:

$$\mathcal{K}(G) = \frac{1}{2m} \sum_{i=2}^{n} \frac{1}{\lambda_i} \sum_{j<k} d_j d_k (U_{j,i} - U_{k,i})^2$$

Resulting expressions for Kemeny's Constant

Additional definitions

For any n and $1 \leq j \leq n$, we define the following vectors:

$$\begin{cases} w_j & = \begin{bmatrix} 0 & 2 & 4 & \dots & 2(j-2) & (j-1)-(j-1)^2 & 0 & \dots & 0 \end{bmatrix} \\ z_j & = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & j & 1 & \dots & 1 \end{bmatrix}, \text{ with } j-1 \text{ zeros} \end{cases}$$

Expression 2 (directly from the code)

Let G be a threshold graph on n vertices. Then:

$$\mathcal{K}(G) = n - 1 - \sum_{j=2}^{n} \frac{c_j}{z_j \cdot c} + \sum_{j=2}^{n} \frac{(w_j \cdot c)(2m - w_j \cdot c)}{2m \cdot j(j-1) \cdot (z_j \cdot c)}$$

The Flip Transformation

Flip Transformation

Let x_1 and x_2 be arbitrary binary sequences and suppose that C is the code $C = x_1 01 x_2 1$. Define the flip transformation as the mapping

$$C = x_1 01 x_2 1 \mapsto x_1 10 x_2 1$$

The inverse transformation works in the expected way: if $C = x_1 10 x_2$, then the $'10'$ block becomes a $'01'$.

Definition

A flip transformation is called a "Braess" flip if it decreases Kemeny's constant.

Remaining Conjectures

Conjecture

We have the following asymptotic behavior of the extremal values of Kemeny's constant:

$$\lim_{n\to\infty} \frac{\max_{|V(G)|=n} \mathcal{K}(\mathcal{G})}{n} = \lim_{n\to\infty} \frac{\min_{|V(G)|=n} \mathcal{K}(\mathcal{G})}{n} = 1$$

Conjecture

Given a threshold graph G, there are no Braess flips at positions $(k, k+1)$ for $k \in (\frac{n}{10}, \frac{9n}{10})$.

Combinatorial Approach

The F Matrix

Motivation

There is a combinatorial definition of Kemeny's constant that we can use to approach the problem. First, we need to introduce the F matrix.

Let G be a connected graph with vertices $v_1, \ldots v_n$.

Spanning 2-Forests

We say a pair of disjoint trees (T_{v_i}, T_{v_j}) is a spanning 2-forest of G if $T_{v_i} \cup T_{v_j}$ spans G and $v_i \in T_{v_i}$ and $v_j \in T_{v_j}$

The F Matrix

Motivation

There is a combinatorial definition of Kemeny's constant that we can use to approach the problem. First, we need to introduce the F matrix.

Let G be a connected graph with vertices $v_1, \ldots v_n$.

Spanning 2-Forests

We say a pair of disjoint trees (T_{v_i}, T_{v_j}) is a spanning 2-forest of G if $T_{v_i} \cup T_{v_j}$ spans G and $v_i \in T_{v_i}$ and $v_j \in T_{v_j}$

F Matrix

We can define the F matrix as $F = (f_{i,j})_{1 \leq i,j \leq n}$, where:

$$f_{i,j} = \# \text{ of spanning 2-forests separating } i \text{ and } j$$

Combinatorial Definition

Last formula for Kemeny's Constant (we promise)

Suppose G is a connected graph on n vertices, d is its degree vector, and τ is the number of spanning trees of G. Then we have the formula:

$$\mathcal{K}(G) = \frac{d^T F d}{4m\tau}$$

Combinatorial Definition

Last formula for Kemeny's Constant (we promise)

Suppose G is a connected graph on n vertices, d is its degree vector, and τ is the number of spanning trees of G. Then we have the formula:

$$\mathcal{K}(G) = \frac{d^T F d}{4m\tau}$$

The Problem

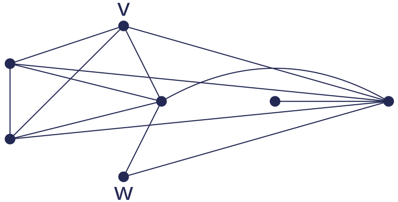
The source of grief in this expression is the F matrix. The rest are easily obtainable directly from the code, so not too difficult to work with.

Theorem

Let G be a threshold graph, and let $x, v, w \in V(G)$ be 3 distinct vertices. Then $f_{x,v} \geq f_{x,w} \iff d_v \leq d_w$.

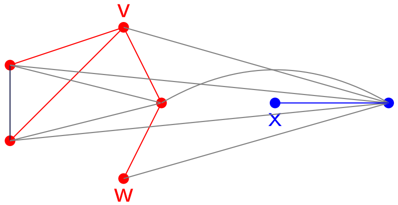
Example

Consider the threshold graph generated by the code $01\underbrace{1}_{\text{v}}\underbrace{0}_{\text{w}}101$ represents the following threshold graph. We have that $f_{x,v} < f_{x,w}$ for all $x \notin \{v, w\}$.



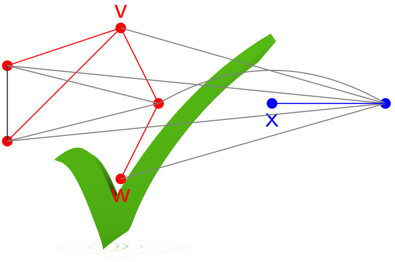
The idea

Consider the following spanning 2-forest separating v and x, where v and w are in the same tree. We find an injective map to the set of spanning 2-forests separating w and x.



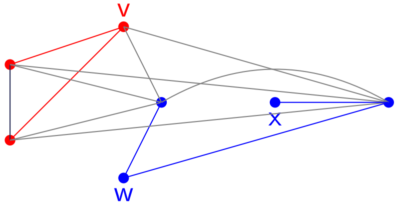
The idea

Consider the following spanning 2-forest separating v and x, where v and w are in the same tree. We find an injective map to the set of spanning 2-forests separating w and x. It's the identity!



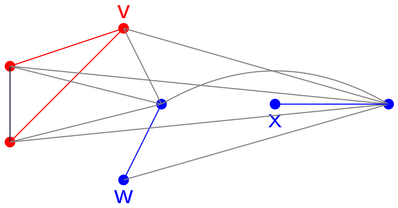
The idea

Now, suppose that x and v are in different trees of the spanning 2-forest.



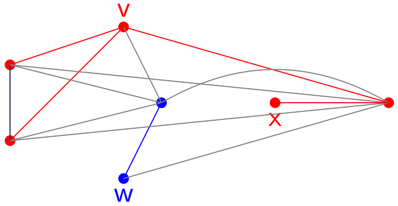
The idea

We construct an injective map to give us a 2-spanning forest that separates w and x.



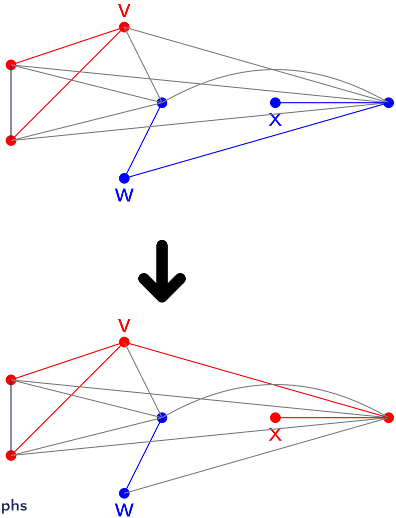
The idea

We construct an injective map to give us a 2-spanning forest that separates w and x.



The idea

We construct an injective map that gives us a 2-spanning forest that separates w and x.



Total Ordering of F

Suppose that we have a code of the form $C = 0^{s_1}1^{t_1}\cdots 0^{s_k}1^{t_k}$ then we have the following ordering for the rows of the F matrix: Let $i \in V(G)$ and denote by $S_\alpha = f_{i,v}$, for $v \in 0^{s_\alpha}, v \neq i$ and $T_\alpha = f_{i,v}$, for $v \in 1^{t_\alpha}, v \neq i$.

Total Ordering of F

Suppose that we have a code of the form $C = 0^{s_1}1^{t_1}\cdots 0^{s_k}1^{t_k}$ then we have the following ordering for the rows of the F matrix: Let $i \in V(G)$ and denote by $S_\alpha = f_{i,v}$, for $v \in 0^{s_\alpha}, v \neq i$ and $T_\alpha = f_{i,v}$, for $v \in 1^{t_\alpha}, v \neq i$.

Ordering Case 1:

$0 = f_{i,i} < T_k \leq T_{k-1} < T_{k-2} < \ldots < T_1 \leq S_1 < S_2 < \ldots < S_{\alpha-1} < S_\alpha < S_{\alpha+1} < \ldots < S_k$

Total Ordering of F

Suppose that we have a code of the form $C = 0^{s_1}1^{t_1}\cdots 0^{s_k}1^{t_k}$ then we have the following ordering for the rows of the F matrix: Let $i \in V(G)$ and denote by $S_\alpha = f_{i,v}$, for $v \in 0^{s_\alpha}, v \neq i$ and $T_\alpha = f_{i,v}$, for $v \in 1^{t_\alpha}, v \neq i$.

Ordering Case 1:

$0 = f_{i,i} < T_k \leq T_{k-1} < T_{k-2} < \ldots < T_1 \leq S_1 < S_2 < \ldots < S_{\alpha-1} < S_\alpha < S_{\alpha+1} < \ldots < S_k$

Ordering Case 2:

$0 = f_{i,i} < T_k \leq T_{k-1} < T_{k-2} < \cdots < T_{\alpha+1} < T_\alpha < T_{\alpha-1} < \ldots < T_1 \leq S_1 < S_2 < \ldots < S_k$

The R Matrix

Now, given a graph G, we can consider the electrical circuit in which each vertex of G is a node and each edge of G is a unit resistor. From here, we can define the R matrix as

$$R_{i,j} = \text{Equivalent resistance between vertices } i \text{ and } j$$

The R Matrix

Now, given a graph G, we can consider the electrical circuit in which each vertex of G is a node and each edge of G is a unit resistor. From here, we can define the R matrix as

$$R_{i,j} = \text{Equivalent resistance between vertices } i \text{ and } j$$

Remark

Let τ denote the total number of spanning trees of G. We have that $F = \tau R$.

Corollary

By ordering F, we also order R.

Explicit form of R

Remark

For a graph G with the Laplacian matrix L and the resistance matrix R, we have:

$$R_{i,j} = (e_i - e_j)^T L(e_i - e_j)$$

Theorem

This allows us to directly compute R from the code using the U matrix:

$$R_{j,k} = \sum_{i=2}^{n} \frac{1}{\lambda_i}(U_{j,i} - U_{k,i})^2$$

Corollaries

The Moment

We define the moment of a vertex v in a graph G as

$$\mu(G, v) = d^T Re_i = \sum_j d_j R_{j,i}$$

where d is the degree matrix and R is the resistance matrix.

Corollary: Ordering of the Moments

For the code $C = 0^{s_1}1^{t_1}0^{s_2}1^{t_2}\ldots0^{s_k}1^{s_k}$, we have the following ordering of the moments:

$$\begin{aligned}\mu(G, 0^{s_k}) > \mu(G, 0^{s_{k-1}}) > \ldots > \mu(G, 0^{s_1}) \\ \geq \mu(G, 1^{t_1}) > \mu(G, 1^{t_2}) \ldots > \mu(G, 1^{t_k})\end{aligned}$$

Corollaries

Accessibility Index

The accessibility index can be defined as

$$\alpha(G, v) = \mu(G, v) - K(G)$$

Corollary: Ordering of Accessibility Indices

$$\alpha(G, v) > \alpha(G, w) \iff \mu(G, v) > \mu(G, w) \iff d_v < d_w$$

Corollaries

Commute Times

We define the commute time as the expected length of a walk that starts at vertex i goes through vertex j and returns to vertex i. We denote this matrix as C.

Corollary: Ordering the Commute Times

Let G be a threshold graph and let $v, w, r \in V(G)$ be 3 distinct vertices. Then we have:

$$C_{v,w} \geq C_{v,r} \iff d_w \leq d_r$$

Remarks

Fruit for Thought

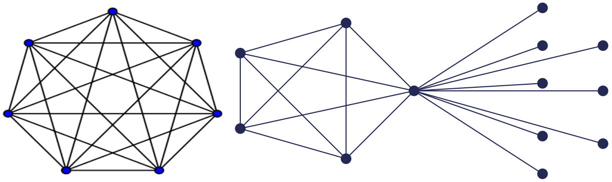
Summary

Kemeny's constant is difficult to work with. We suspect the maximizing graph for a given n is a pineapple graph with approximately $\sqrt{2n}$ ones. We have developed combinatorial and linear algebra tools to examine these extremal cases.

Fruit for Thought

Summary

Kemeny's constant is difficult to work with. We suspect the maximizing graph for a given n is a pineapple graph with approximately $\sqrt{2n}$ ones. We have developed combinatorial and linear algebra tools to examine these extremal cases.



Fruit for Thought

Summary

Kemeny's constant is difficult to work with. We suspect the maximizing graph for a given n is a pineapple graph with approximately $\sqrt{2n}$ ones. We have developed combinatorial and linear algebra tools to examine these extremal cases.

References

References I

Anđelić, Milica and Slobodan K Simić (2010). "Some notes on the threshold graphs". In: *Discrete mathematics* 310.17-18, pp. 2241–2248.

Banerjee, Anirban and Ranjit Mehatari (2017). "On the normalized spectrum of threshold graphs". In: *Linear Algebra and its Applications* 530, pp. 288–304.

Breen, Jane, Emanuele Crisostomi, and Sooyeong Kim (2022). "Kemeny's constant for a graph with bridges". In: *Discrete Applied Mathematics* 322, pp. 20–35.

Brouwer, Andries E and Willem H Haemers (2011). *Spectra of graphs*. Springer Science & Business Media.

Butler, Steve, Fan Chung, et al. (2006). "Spectral graph theory". In: *Handbook of linear algebra*, pp. 24–25.

Diagonalization by a unitary similarity transformation (2011). url: `http://scipp.ucsc.edu/~haber/ph116A/diag_11.pdf`.

References II

- Faught, Nolan, Mark Kempton, and Adam Knudson (2022). "A 1-separation formula for the graph Kemeny constant and Braess edges". In: *Journal of Mathematical Chemistry* 60.1, pp. 49–69.
- Hammer, Peter L and Alexander K Kelmans (1996). "Laplacian spectra and spanning trees of threshold graphs". In: *Discrete Applied Mathematics* 65.1-3, pp. 255–273.
- Hu, Yuxiang and Steve Kirkland (2019). "Complete multipartite graphs and Braess edges". In: *Linear Algebra and its Applications* 579, pp. 284–301.
- Kirkland, Steve (2016). "Random walk centrality and a partition of Kemeny's constant". In: *Czechoslovak Mathematical Journal* 66, pp. 757–775.
- Mishra, Ankit, Ranveer Singh, and Sarika Jalan (2022). "On the second largest eigenvalue of networks". In: *Applied Network Science* 7.1, p. 47.
- Sheskin, Theodore J (1995). "Computing mean first passage times for a Markov chain". In: *International Journal of Mathematical Education in Science and Technology* 26.5, pp. 729–735.