

The edit distance on graphs, Part I

Ryan Martin

Iowa State University

5th Lake Michigan Workshop on Combinatorics and Graph Theory
Notre Dame University

Survey: [The edit distance in graphs: methods, results and generalizations](#), *Recent Trends in Combinatorics*, 31–62, IMA Vol. Math. Appl., **159**, Springer, Cham, 2016.

- 1 Metric Spaces
- 2 Extremal Graph Theory
- 3 The Edit Distance Problem
- 4 Hereditary Properties
- 5 The Edit Distance Function
- 6 Binary Chromatic Number
- 7 Results

Stability

Turán's theorem says that the K_{p+1} -free n -vertex graph with the most edges is the Turán graph, $T_{n,p}$.

Theorem (Turán, 1941)

Suppose that G is an K_{p+1} -free n -vertex graph. Then $e(G) \leq e(T_{n,p})$.

Remark

$T_{n,p}$ is a balanced n -vertex p -partite graph and $e(T_{n,p}) \approx \frac{p-1}{p} \binom{n}{2}$.

Remark

$T_{n,p}$ is a balanced n -vertex p -partite graph and $e(T_{n,p}) \approx \frac{p-1}{p} \binom{n}{2}$.

Although Simonovits had one of the first *stability theorems* in 1966, a very precise version is due to Füredi.

Theorem (Füredi, 2015)

Suppose that G is an K_{p+1} -free n -vertex graph with $e(G) \geq e(T_{n,p}) - t$. There is a complete p -partite graph G' with $V(G') = V(G)$, such that

$$|E(G) \Delta E(G')| \leq 3t.$$

Remark

$T_{n,p}$ is a balanced n -vertex p -partite graph and $e(T_{n,p}) \approx \frac{p-1}{p} \binom{n}{2}$.

Although Simonovits had one of the first *stability theorems* in 1966, a very precise version is due to Füredi.

Theorem (Füredi, 2015)

Suppose that G is an K_{p+1} -free n -vertex graph with $e(G) \geq e(T_{n,p}) - t$. There is a complete p -partite graph G' with $V(G') = V(G)$, such that

$$|E(G) \Delta E(G')| \leq 3t.$$

The **EDIT DISTANCE** between G and G' is $|E(G) \Delta E(G')|$.

Remark

$T_{n,p}$ is a balanced n -vertex p -partite graph and $e(T_{n,p}) \approx \frac{p-1}{p} \binom{n}{2}$.

Although Simonovits had one of the first *stability theorems* in 1966, a very precise version is due to Füredi.

Theorem (Füredi, 2015)

Suppose that G is an K_{p+1} -free n -vertex graph with $e(G) \geq e(T_{n,p}) - t$. There is a complete p -partite graph G' with $V(G') = V(G)$, such that

$$|E(G) \Delta E(G')| \leq 3t.$$

The **EDIT DISTANCE** between G and G' is $|E(G) \Delta E(G')|$. The **NORMALIZED EDIT DISTANCE** between G and G' is

$$|E(G) \Delta E(G')| / \binom{n}{2}.$$

For every n , let us consider the following finite metric space \mathcal{G}_n :

- The elements are graphs on the *labeled* vertex set $V = \{1, 2, \dots, n\}$.

For every n , let us consider the following finite metric space \mathcal{G}_n :

- The elements are graphs on the *labeled* vertex set $V = \{1, 2, \dots, n\}$.
- The distance between graphs G', G'' is the normalized edit distance.

$$\text{dist}(G', G'') = |E(G') \Delta E(G'')| / \binom{n}{2}$$

Metric spaces

For every n , let us consider the following finite metric space \mathcal{G}_n :

- The elements are graphs on the *labeled* vertex set $V = \{1, 2, \dots, n\}$.
- The distance between graphs G', G'' is the normalized edit distance.

$$\text{dist}(G', G'') = |E(G') \Delta E(G'')| / \binom{n}{2}$$

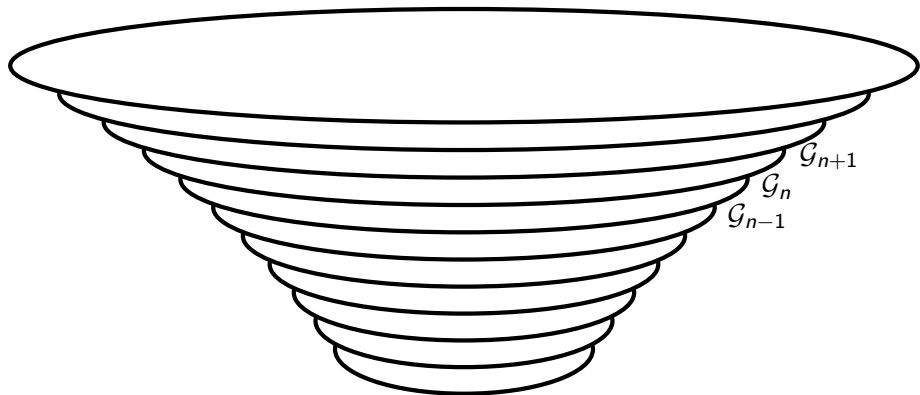
Metric Space

- **Identity:** $\text{dist}(G', G'') = 0 \Leftrightarrow G' = G''$
- **Symmetry:** $\text{dist}(G', G'') = \text{dist}(G'', G')$
- **Triangle:** $\text{dist}(G', G''') \leq \text{dist}(G', G'') + \text{dist}(G'', G''')$ because

$$|A \Delta C| \leq |A \Delta B| + |B \Delta C|.$$

Sequence of metric spaces

So we have a sequence of metric spaces $\{\mathcal{G}_n\}$.



Examples

Fix some n and let K_n be the complete graph and E_n be the empty graph:

$$\text{dist}(K_n, E_n) =$$

Examples

Fix some n and let K_n be the complete graph and E_n be the empty graph:

$$\text{dist}(K_n, E_n) = 1.$$

Examples

Fix some n and let K_n be the complete graph and E_n be the empty graph:

$$\text{dist}(K_n, E_n) = 1.$$

Question

What is $\text{dist}(C_5, C_5)$?

Examples

Fix some n and let K_n be the complete graph and E_n be the empty graph:

$$\text{dist}(K_n, E_n) = 1.$$

Question

What is $\text{dist}(C_5, C_5)$?

Answer

0 or $2/5$ or $3/5$ or 1.

Examples

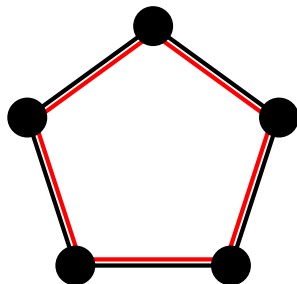
Question

What is $\text{dist}(C_5, C_5)$?

Answer

0 or $2/5$ or $3/5$ or 1 .

Remember that the distance is between graphs on a *labeled* vertex set.



Examples

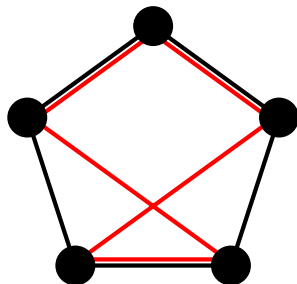
Question

What is $\text{dist}(C_5, C_5)$?

Answer

0 or $2/5$ or $3/5$ or 1.

Remember that the distance is between graphs on a *labeled* vertex set.



Examples

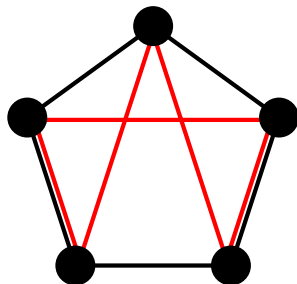
Question

What is $\text{dist}(C_5, C_5)$?

Answer

0 or $2/5$ or $3/5$ or 1.

Remember that the distance is between graphs on a *labeled* vertex set.



Examples

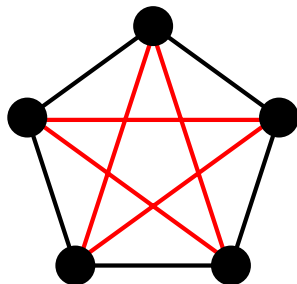
Question

What is $\text{dist}(C_5, C_5)$?

Answer

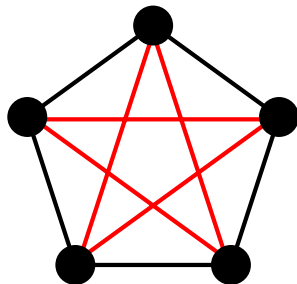
0 or $2/5$ or $3/5$ or 1 .

Remember that the distance is between graphs on a *labeled* vertex set.



Examples

Remember that the distance is between graphs on a *labeled* vertex set.



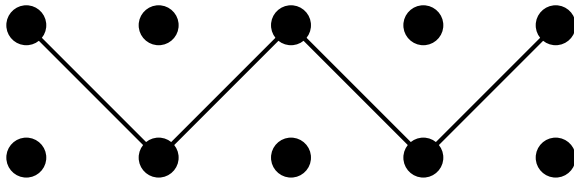
- We want the distance between graphs to be the distance between isomorphism classes.
- The distance between isomorphism classes of G and G' is the same as the distance between any copy of G and the isomorphism class of G' .
- So, we will consider the distance between a graph and a large set of graphs, which is closed under isomorphism.

A question in biology

Maria Axenovich and I were asked a question:

Question

For every $n \times n$ bipartite graph, is it possible to add and delete edges at most $n^2/3$ times so that the result has no induced copy of P_5 , the path on 5 vertices?

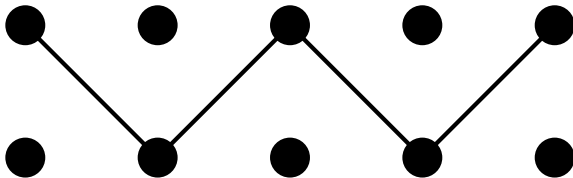


A question in biology

Maria Axenovich and I were asked a question:

Question

For every $n \times n$ bipartite graph, is it possible to add and delete edges at most $n^2/3$ times so that the result has no induced copy of P_5 , the path on 5 vertices?



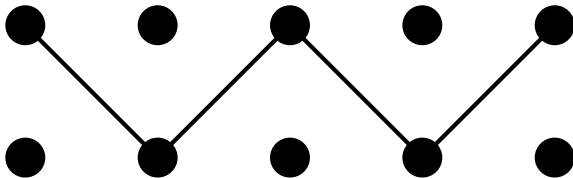
The motivation was from some work of Chen, Eulenstein, Fernández-Baca, and Sanderson on consensus trees in evolutionary biology.

A question in biology

Maria Axenovich and I were asked a question:

Question

For every $n \times n$ bipartite graph, is it possible to add and delete edges at most $n^2/3$ times so that the result has no induced copy of P_5 , the path on 5 vertices?



The motivation was from some work of Chen, Eulenstein, Fernández-Baca, and Sanderson on consensus trees in evolutionary biology.

Answer

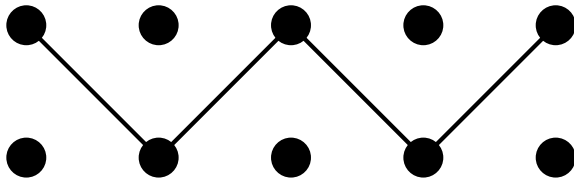
No.

A question in biology

Maria Axenovich and I were asked a question:

Question

For every $n \times n$ bipartite graph, is it possible to add and delete edges at most $n^2/3$ times so that the result has no induced copy of P_5 , the path on 5 vertices?



The motivation was from some work of Chen, Eulenstein, Fernández-Baca, and Sanderson on consensus trees in evolutionary biology.

Answer

No.

But we'll get back to that later.

If an n -vertex graph G on a labeled vertex set is modified by
edge-additions plus edge-deletions

to create the graph G' , then the number of operations required to do so is

$$|E(G) \Delta E(G')| =$$

If an n -vertex graph G on a labeled vertex set is modified by
edge-additions plus edge-deletions

to create the graph G' , then the number of operations required to do so is

$$|E(G) \Delta E(G')| = \binom{n}{2} \cdot \text{dist}(G, G').$$

The proportion of edge-changes to get from G to G' is $\text{dist}(G, G')$.

If an n -vertex graph G on a labeled vertex set is modified by
edge-additions plus edge-deletions

to create the graph G' , then the number of operations required to do so is

$$|E(G) \Delta E(G')| = \binom{n}{2} \cdot \text{dist}(G, G').$$

The proportion of edge-changes to get from G to G' is $\text{dist}(G, G')$.

Now suppose we want to get rid of some kind of induced subgraph from a graph G .

A classical problem

Problem

Among all n -vertex graphs, what is the maximum proportion of edge-additions and edge-deletions necessary so that the resulting graph has no triangles?

A classical problem

Problem

Among all n -vertex graphs, what is the maximum proportion of edge-additions and edge-deletions necessary so that the resulting graph has no triangles?

Theorem (Mantel, 1907)

If an n -vertex graph G' has no triangles, then the number of edges H has is at most $\lfloor n^2/4 \rfloor$.

This bound is only achieved if G' is complete bipartite. E.g., $G' = K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

A classical problem

Theorem (Mantel, 1907)

If an n -vertex graph G' has no triangles, then the number of edges H has is at most $\lfloor n^2/4 \rfloor$.

This bound is only achieved if G' is complete bipartite. E.g., $G' = K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

The graph achieving the maximum distance is the complete graph, K_n and

$$\begin{aligned} \text{dist}(K_n, K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}) &= \frac{\binom{n}{2} - \lfloor \frac{n^2}{4} \rfloor}{\binom{n}{2}} \\ &= \frac{1}{2} - \frac{1}{4\lfloor n/2 \rfloor - 2}. \end{aligned}$$

A classical problem

Theorem (Mantel, 1907)

If an n -vertex graph G' has no triangles, then the number of edges H has is at most $\lfloor n^2/4 \rfloor$.

This bound is only achieved if G' is complete bipartite. E.g., $G' = K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

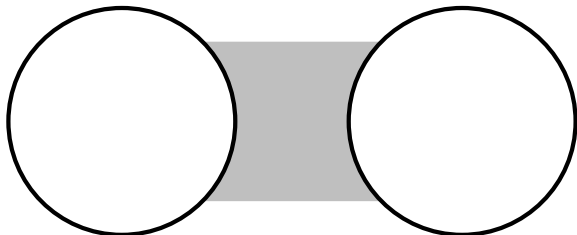
The graph achieving the maximum distance is the complete graph, K_n and

$$\begin{aligned} \text{dist} \left(K_n, K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil} \right) &= \frac{\binom{n}{2} - \lfloor \frac{n^2}{4} \rfloor}{\binom{n}{2}} \\ &= \frac{1}{2} - o(1). \end{aligned}$$

A classical problem

The graph achieving the maximum distance is the complete graph, K_n and

$$\begin{aligned}\text{dist} (K_n, K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}) &= \frac{\binom{n}{2} - \lfloor \frac{n^2}{4} \rfloor}{\binom{n}{2}} \\ &= \frac{1}{2} - o(1).\end{aligned}$$



A classical problem

The graph achieving the maximum distance is the complete graph, K_n and

$$\begin{aligned}\text{dist}(K_n, K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}) &= \frac{\binom{n}{2} - \lfloor \frac{n^2}{4} \rfloor}{\binom{n}{2}} \\ &= \frac{1}{2} - o(1).\end{aligned}$$

Algorithm

For any graph G , we can eliminate triangles by

- partitioning $V(G)$ into 2 parts, and
- deleting all the edges inside each part.

Cliques

Instead of triangles, we can generalize the previous analysis to forbid copies of K_{p+1} , $p \geq 2$.

Theorem (Turán, 1941)

If an n -vertex graph G' has no copy of K_{p+1} , then

$$e(G') \leq e(T_{n,p}) = \left(\frac{p-1}{p} - o(1) \right) \frac{n^2}{2}.$$

The maximum is achieved by the Turán graph $T(n, p)$, a balanced complete p -partite graph.

Cliques

Instead of triangles, we can generalize the previous analysis to forbid copies of K_{p+1} , $p \geq 2$.

Theorem (Turán, 1941)

If an n -vertex graph G' has no copy of K_{p+1} , then

$$e(G') \leq e(T_{n,p}) = \left(\frac{p-1}{p} - o(1) \right) \frac{n^2}{2}.$$

The maximum is achieved by the Turán graph $T(n, p)$, a balanced complete p -partite graph.

The graph achieving the maximum distance is the complete graph, K_n and

$$\text{dist}(K_n, T_{n,p}) = \frac{1}{p} - o(1).$$

The graph achieving the maximum distance is the complete graph, K_n and

$$\text{dist}(K_n, T_{n,p}) = \frac{1}{p} - o(1).$$

Algorithm

For any graph G , we can eliminate K_{p+1} -s by

- partitioning $V(G)$ into p parts, and
- deleting all the edges inside each part.

Graph properties

Given: An n -vertex graph G on a labeled vertex set and graph property \mathcal{P} .
A **GRAPH PROPERTY** is a set of graphs.

Definition

The **EDIT DISTANCE FROM G TO \mathcal{P}**

$$\text{dist}(G, \mathcal{P}) = \min \{ \text{dist}(G, G') : G' \in \mathcal{P} \}$$

is the least edit distance of G to a graph in \mathcal{P} .

Graph properties

Given: An n -vertex graph G on a labeled vertex set and graph property \mathcal{P} .
A **GRAPH PROPERTY** is a set of graphs.

Definition

The **EDIT DISTANCE FROM G TO \mathcal{P}**

$$\text{dist}(G, \mathcal{P}) = \min \{ \text{dist}(G, G') : G' \in \mathcal{P} \}$$

is the least edit distance of G to a graph in \mathcal{P} .

That is, it is the minimum proportion of edge-additions plus edge-deletions to transform G into a member of \mathcal{P} .

Graph properties

Given: An n -vertex graph G on a labeled vertex set and graph property \mathcal{P} .
A **GRAPH PROPERTY** is a set of graphs.

Definition

The **EDIT DISTANCE FROM G TO \mathcal{P}**

$$\begin{aligned}\text{dist}(G, \mathcal{P}) &= \min \{ \text{dist}(G, G') : G' \in \mathcal{P} \} \\ &= \min \{ |E(G) \Delta E(G')| / \binom{n}{2} : G' \in \mathcal{P} \}\end{aligned}$$

is the least edit distance of G to a graph in \mathcal{P} .

That is, it is the minimum proportion of edge-additions plus edge-deletions to transform G into a member of \mathcal{P} .

Extremal Edit Distance

Given: A natural number n and a graph property \mathcal{P} .

Definition

The **EDIT DISTANCE FROM \mathcal{P}**

$$\text{dist}(n, \mathcal{P}) = \max \{ \text{dist}(G, \mathcal{P}) : |V(G)| = n \}$$

is the maximum edit distance of an n -vertex graph to a graph in \mathcal{P} .

Extremal Edit Distance

Given: A natural number n and a graph property \mathcal{P} .

Definition

The **EDIT DISTANCE FROM \mathcal{P}**

$$\text{dist}(n, \mathcal{P}) = \max \{ \text{dist}(G, \mathcal{P}) : |V(G)| = n \}$$

is the maximum edit distance of an n -vertex graph to a graph in \mathcal{P} .

That is:

- the maximum, over all n -vertex graphs, G ,
- of the minimum proportion of edge-additions plus edge-deletions
- to transform G into a member of \mathcal{P} .

Extremal Edit Distance

Given: A natural number n and a hereditary graph property \mathcal{H} .

Definition

The **EDIT DISTANCE FROM \mathcal{H}**

$$\text{dist}(n, \mathcal{H}) = \max \{ \text{dist}(G, \mathcal{H}) : |V(G)| = n \}$$

is the maximum edit distance of an n -vertex graph to a graph in \mathcal{H} .

That is:

- the maximum, over all n -vertex graphs, G ,
- of the minimum proportion of edge-additions plus edge-deletions
- to transform G into a member of \mathcal{H} .

A **HEREDITARY PROPERTY** is one that is preserved under isomorphism and vertex-deletion.

Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.

Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.

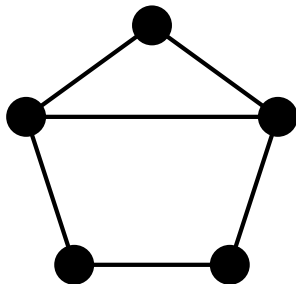
I.e., if $G \in \mathcal{H}$, then every induced subgraph of G is in \mathcal{H} .

Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.

I.e., if $G \in \mathcal{H}$, then every induced subgraph of G is in \mathcal{H} .

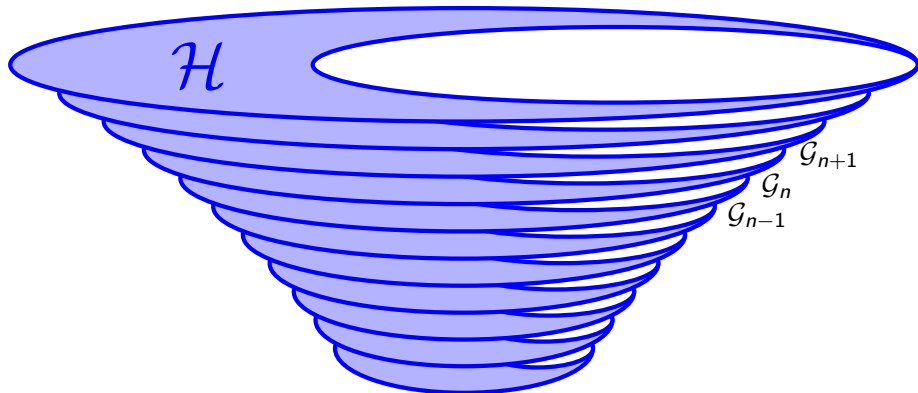


A 5-cycle as a subgraph, but no *induced* 5-cycle.

Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.



Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.

I.e., if $G \in \mathcal{H}$, then every induced subgraph of G is in \mathcal{H} .

Examples:

- Planar graphs.
- Line graphs of bipartite graphs.
- Chordal graphs: graphs with no chordless cycle of length ≥ 4 .
- Perfect graphs: $\chi = \omega$ for all induced subgraphs.
- $\text{Forb}(H)$

Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.

I.e., if $G \in \mathcal{H}$, then every induced subgraph of G is in \mathcal{H} .

Examples:

- Planar graphs.
- Line graphs of bipartite graphs.
- Chordal graphs: graphs with no chordless cycle of length ≥ 4 .
- Perfect graphs: $\chi = \omega$ for all induced subgraphs.
- $\text{Forb}(H)$: graphs with no *induced* copy of H .

Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.

I.e., if $G \in \mathcal{H}$, then every induced subgraph of G is in \mathcal{H} .

Examples:

- Planar graphs.
- Line graphs of bipartite graphs.
- Chordal graphs: graphs with no chordless cycle of length ≥ 4 .
- Perfect graphs: $\chi = \omega$ for all induced subgraphs.
- $\text{Forb}(H)$: graphs with no *induced* copy of H .
($\text{Forb}(H)$ is a **PRINCIPAL HEREDITARY PROPERTY**.)

Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.

i.e., if $G \in \mathcal{H}$, then every induced subgraph of G is in \mathcal{H} .

Examples:

- Planar graphs.
- Line graphs of bipartite graphs.
- Chordal graphs: graphs with no chordless cycle of length ≥ 4 .
- Perfect graphs: $\chi = \omega$ for all induced subgraphs.
- $\text{Forb}(H)$: graphs with no *induced* copy of H .
($\text{Forb}(H)$ is a **PRINCIPAL HEREDITARY PROPERTY**.)

For the rest of this talk, all of our hereditary properties are principal; i.e.,

$$\mathcal{H} = \text{Forb}(H), \text{ for some graph } H.$$

Hereditary properties

Definition

A **HEREDITARY PROPERTY**, \mathcal{H} , of graphs is one that holds under isomorphism and the deletion of vertices.

I.e., if $G \in \mathcal{H}$, then every induced subgraph of G is in \mathcal{H} .

Examples:

- $\text{Forb}(H)$: graphs with no *induced* copy of H .
($\text{Forb}(H)$ is a **PRINCIPAL HEREDITARY PROPERTY**.)

For the rest of this talk, all of our hereditary properties are principal; i.e.,

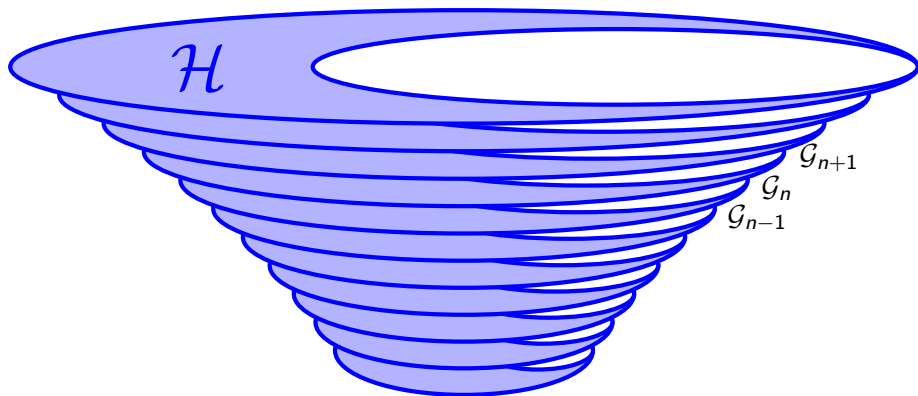
$$\mathcal{H} = \text{Forb}(H), \text{ for some graph } H.$$

In fact, every hereditary property is of the form $\bigcap_{H \in \mathcal{F}(\mathcal{H})} \text{Forb}(H)$.

Understanding d^*

Problem

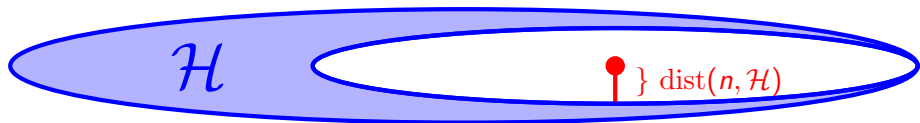
Compute $\text{dist}(n, \mathcal{H})$ for a hereditary property \mathcal{H} .



Understanding d^*

Problem

Compute $\text{dist}(n, \mathcal{H})$ for a hereditary property \mathcal{H} .



Problem

Compute $\text{dist}(n, \mathcal{H})$ for a hereditary property \mathcal{H} .

Let $G(n, p)$ denote the Erdős-Rényi random graph:

- n vertices,
- each edge is present, independently, with probability p .

Understanding d^*

Problem

Compute $\text{dist}(n, \mathcal{H})$ for a hereditary property \mathcal{H} .

Let $G(n, p)$ denote the Erdős-Rényi random graph:

- n vertices,
- each edge is present, independently, with probability p .

Theorem (Alon-Stav, 2008)

For every hereditary property, \mathcal{H} , the following limit exists

$$d_{\mathcal{H}}^* = \lim_{n \rightarrow \infty} \text{dist}(n, \mathcal{H}).$$

Furthermore, there exists a $p^* = p_{\mathcal{H}}^* \in [0, 1]$ such that

$$d_{\mathcal{H}}^* = \lim_{n \rightarrow \infty} \mathbb{E}[\text{dist}(G(n, p^*), \mathcal{H})].$$

Understanding d^*

Theorem (Alon-Stav, 2008)

For every hereditary property, \mathcal{H} , the following limit exists

$$d_{\mathcal{H}}^* = \lim_{n \rightarrow \infty} \text{dist}(n, \mathcal{H}).$$

Furthermore, there exists a $p^* = p_{\mathcal{H}}^* \in [0, 1]$ such that

$$d_{\mathcal{H}}^* = \lim_{n \rightarrow \infty} \mathbb{E}[\text{dist}(G(n, p^*), \mathcal{H})].$$

Theorem (Concentration)

Let $\mu = \mathbb{E}[\text{dist}(G(n, p), \mathcal{H})]$. Then for every $\epsilon > 0$,

$$\Pr\{|\text{dist}(G(n, p), \mathcal{H}) - \mu| \geq \epsilon\} \rightarrow 0.$$

So, $G(n, p) \sim \mu$ *asymptotically almost surely (a.a.s.)*.

Theorem (Alon-Stav, 2008)

For every hereditary property, \mathcal{H} , the following limit exists

$$d_{\mathcal{H}}^* = \lim_{n \rightarrow \infty} \text{dist}(n, \mathcal{H}).$$

Furthermore, there exists a $p^* = p_{\mathcal{H}}^* \in [0, 1]$ such that

$$d_{\mathcal{H}}^* = \lim_{n \rightarrow \infty} \mathbb{E}[\text{dist}(G(n, p^*), \mathcal{H})].$$

So, given \mathcal{H} , what is $p_{\mathcal{H}}^*$?

Generalizing helps!

The problem of finding $p_{\mathcal{H}}^*$ is difficult.
It's made easier by not finding it directly.

Generalizing helps!

The problem of finding $p_{\mathcal{H}}^*$ is difficult.
It's made easier by not finding it directly.

Theorem (Balogh-M., 2008)

For every hereditary property, \mathcal{H} , and every $p \in [0, 1]$, if we define

$$\text{ed}_{\mathcal{H}}(p) = \lim_{n \rightarrow \infty} \mathbb{E} [\text{dist}(G(n, p), \mathcal{H})]$$

then the limit exists and

$$\text{ed}_{\mathcal{H}}(p) = \lim_{n \rightarrow \infty} \max \left\{ \text{dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p \binom{n}{2} \rfloor \right\}.$$

Generalizing helps!

The problem of finding $p_{\mathcal{H}}^*$ is difficult.
It's made easier by not finding it directly.

Theorem (Balogh-M., 2008)

For every hereditary property, \mathcal{H} , and every $p \in [0, 1]$, if we define

$$\text{ed}_{\mathcal{H}}(p) = \lim_{n \rightarrow \infty} \mathbb{E} [\text{dist}(G(n, p), \mathcal{H})]$$

then the limit exists and

$$\text{ed}_{\mathcal{H}}(p) = \lim_{n \rightarrow \infty} \max \left\{ \text{dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p \binom{n}{2} \rfloor \right\}.$$

Roughly, any density- p graph is at most as hard to edit as the same density random graph.

The Edit Distance Function

Properties of $\text{ed}_{\mathcal{H}}(p)$, $\mathcal{H} = \text{Forb}(H)$

The Edit Distance Function

Properties of $\text{ed}_{\mathcal{H}}(p)$, $\mathcal{H} = \text{Forb}(H)$

- Continuous and concave down.

The Edit Distance Function

Properties of $\text{ed}_{\mathcal{H}}(p)$, $\mathcal{H} = \text{Forb}(H)$

- Continuous and concave down.
- Achieves its maximum (p^*, d^*) for some $p^* \in [0, 1]$.

The Edit Distance Function

Properties of $\text{ed}_{\mathcal{H}}(p)$, $\mathcal{H} = \text{Forb}(H)$

- Continuous and concave down.
- Achieves its maximum (p^*, d^*) for some $p^* \in [0, 1]$.
- $\text{ed}_{\mathcal{H}}(0)$, $\text{ed}_{\mathcal{H}}(1/2)$, and $\text{ed}_{\mathcal{H}}(1)$ are each computable.

The Edit Distance Function

Properties of $\text{ed}_{\mathcal{H}}(p)$, $\mathcal{H} = \text{Forb}(H)$

- Continuous and concave down.
- Achieves its maximum (p^*, d^*) for some $p^* \in [0, 1]$.
- $\text{ed}_{\mathcal{H}}(0)$, $\text{ed}_{\mathcal{H}}(1/2)$, and $\text{ed}_{\mathcal{H}}(1)$ are each computable.
- If H is neither complete nor empty, then $\text{ed}_{\mathcal{H}}(0) = \text{ed}_{\mathcal{H}}(1) = 0$.

The Edit Distance Function

Properties of $\text{ed}_{\mathcal{H}}(p)$, $\mathcal{H} = \text{Forb}(H)$

- Continuous and concave down.
- Achieves its maximum (p^*, d^*) for some $p^* \in [0, 1]$.
- $\text{ed}_{\mathcal{H}}(0)$, $\text{ed}_{\mathcal{H}}(1/2)$, and $\text{ed}_{\mathcal{H}}(1)$ are each computable.
- If H is neither complete nor empty, then $\text{ed}_{\mathcal{H}}(0) = \text{ed}_{\mathcal{H}}(1) = 0$.
- For any rational $r \in [0, 1]$, there is an H , such that $p_{\text{Forb}(H)}^* = r$.

Theorem (Balogh-M., 2008)

For $a \geq 1$ and $b \geq 0$, let $K_a + E_b$ denote a clique of a vertices, together with b independent vertices. Then,

$$p_{\text{Forb}(K_a + E_b)}^* = \frac{a - 1}{a + b - 1} \quad d_{\text{Forb}(K_a + E_b)}^* = \frac{1}{a + b - 1}.$$

The Edit Distance Function

Properties of $\text{ed}_{\mathcal{H}}(p)$, $\mathcal{H} = \text{Forb}(H)$

- Continuous and concave down.
- Achieves its maximum (p^*, d^*) for some $p^* \in [0, 1]$.
- $\text{ed}_{\mathcal{H}}(0)$, $\text{ed}_{\mathcal{H}}(1/2)$, and $\text{ed}_{\mathcal{H}}(1)$ are each computable.
- If H is neither complete nor empty, then $\text{ed}_{\mathcal{H}}(0) = \text{ed}_{\mathcal{H}}(1) = 0$.
- For any rational $r \in [0, 1]$, there is an H , such that $p_{\text{Forb}(H)}^* = r$.
- There is an H , such that $p_{\text{Forb}(H)}^* = \sqrt{2} - 1$.

Theorem (Balogh-M., 2008)

$$p_{\text{Forb}(K_{3,3})}^* = \sqrt{2} - 1 \quad d_{\text{Forb}(K_{3,3})}^* = 3 - 2\sqrt{2}.$$

What is $\text{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \text{Forb}(H)$ and let $r, s \in \mathbb{N}$ have the property that $V(H)$ cannot be partitioned into a set of r independent sets and s cliques.

But $V(H)$ can be partitioned into *any* combination of $r + s + 1$ independent sets and cliques.

What is $\text{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \text{Forb}(H)$ and let $r, s \in \mathbb{N}$ have the property that $V(H)$ cannot be partitioned into a set of r independent sets and s cliques.

But $V(H)$ can be partitioned into *any* combination of $r + s + 1$ independent sets and cliques.

We call $r + s + 1$ **THE BINARY CHROMATIC NUMBER**, $\chi_B(H)$.

What is $\text{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \text{Forb}(H)$ and let $r, s \in \mathbb{N}$ have the property that $V(H)$ cannot be partitioned into a set of r independent sets and s cliques.

But $V(H)$ can be partitioned into *any* combination of $r + s + 1$ independent sets and cliques.

We call $r + s + 1$ **THE BINARY CHROMATIC NUMBER**, $\chi_B(H)$.

But the parameter was used before.

- $r + s = \tau(\mathcal{H})$ [Prömel-Steger (1992)]
- $r + s = r(\mathcal{H})$, **COLOURING NUMBER** [Bollobás-Thomason (1995)]

“Binary” is because the idea can be generalized to multicolored complete graphs.

There is a generalization of this parameter to all hereditary properties \mathcal{H} .

What is $\text{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \text{Forb}(H)$ and let $r, s \in \mathbb{N}$ have the property that $V(H)$ cannot be partitioned into a set of r independent sets and s cliques.

But $V(H)$ can be partitioned into *any* combination of $r + s + 1$ independent sets and cliques.

We call $r + s + 1$ **THE BINARY CHROMATIC NUMBER**, $\chi_B(H)$.

Theorem (Prömel-Steger, 1993)

Let H be a graph.

$$|\text{Forb}(H)| = 2^{\left(1 - \frac{1}{\chi_B(H)-1}\right) \binom{n}{2} + o(n^2)}.$$

What is $\text{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \text{Forb}(H)$ and let $r, s \in \mathbb{N}$ have the property that $V(H)$ cannot be partitioned into a set of r independent sets and s cliques.

But $V(H)$ can be partitioned into *any* combination of $r + s + 1$ independent sets and cliques.

We call $r + s + 1$ **THE BINARY CHROMATIC NUMBER**, $\chi_B(H)$.

Theorem (Prömel-Steger, 1993; Bollobás-Thomason, 1995)

Let \mathcal{H} be a hereditary property.

$$|\mathcal{H}| = 2^{\left(1 - \frac{1}{\chi_B(\mathcal{H}) - 1}\right) \binom{n}{2} + o(n^2)}.$$

$\chi_B(\mathcal{H})$ was generalized to all hereditary properties by Bollobás and Thomason (1995).

What is $\text{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \text{Forb}(H)$ and let $r, s \in \mathbb{N}$ have the property that $V(H)$ cannot be partitioned into a set of r independent sets and s cliques.

But $V(H)$ can be partitioned into *any* combination of $r + s + 1$ independent sets and cliques.

We call $r + s + 1$ **THE BINARY CHROMATIC NUMBER**, $\chi_B(H)$.

Theorem (Axenovich-Kézdy-M., 2008)

Let $\mathcal{H} = \text{Forb}(H)$ for some fixed graph H which has binary chromatic number $\chi_B(H)$,

$$\text{ed}_{\mathcal{H}}(1/2) = \frac{1}{2(\chi_B(H) - 1)} \leq d_{\mathcal{H}}^*.$$

What is $\text{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \text{Forb}(H)$ and let $r, s \in \mathbb{N}$ have the property that $V(H)$ cannot be partitioned into a set of r independent sets and s cliques.

But $V(H)$ can be partitioned into *any* combination of $r + s + 1$ independent sets and cliques.

We call $r + s + 1$ **THE BINARY CHROMATIC NUMBER**, $\chi_B(H)$.

Theorem (Axenovich-Kézdy-M., 2008)

Let $\mathcal{H} = \text{Forb}(H)$ for some fixed graph H which has binary chromatic number $\chi_B(H)$,

$$\text{ed}_{\mathcal{H}}(1/2) = \frac{1}{2(\chi_B(H) - 1)} \leq d_{\mathcal{H}}^*.$$

Moreover, equality holds if H is self-complementary.

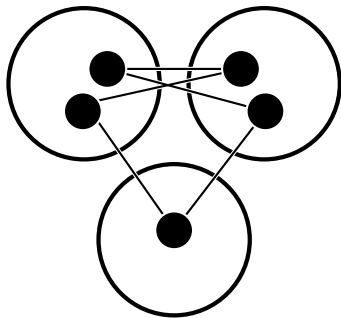
An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

Let us consider all (r, s) such that $r + s + 1 = 3$:

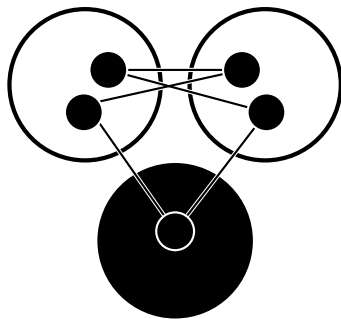


The 5-cycle can be partitioned into
3 independent sets and 0 cliques.

An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

Let us consider all (r, s) such that $r + s + 1 = 3$:

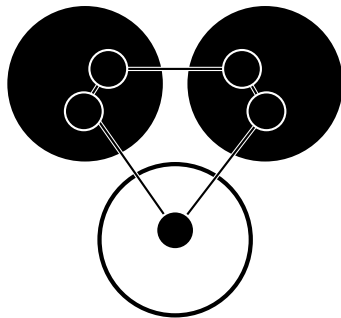


The 5-cycle can be partitioned into
2 independent sets and 1 cliques.

An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

Let us consider all (r, s) such that $r + s + 1 = 3$:

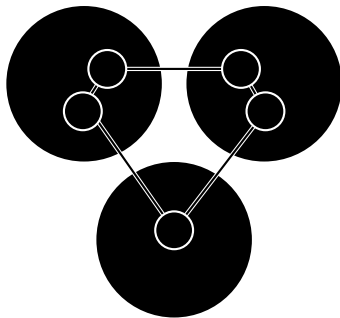


The 5-cycle can be partitioned into
1 independent sets and 2 cliques.

An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

Let us consider all (r, s) such that $r + s + 1 = 3$:

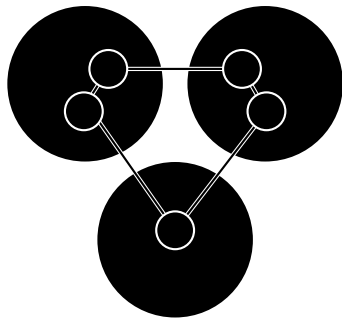


The 5-cycle can be partitioned into
0 independent sets and 3 cliques.

An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

Let us consider all (r, s) such that $r + s + 1 = 3$:



The 5-cycle can be partitioned into
0 independent sets and 3 cliques.

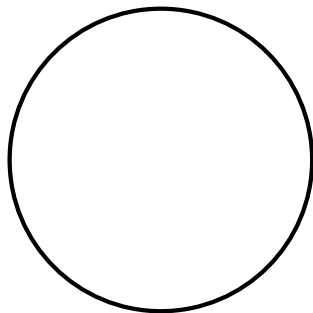
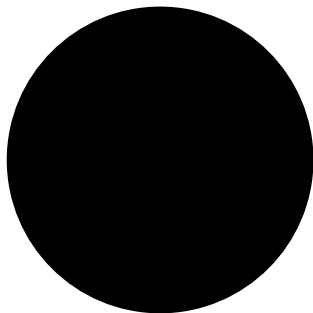
So, $\chi_B(C_5) \leq 3$.

An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

So, $\chi_B(C_5) \leq 3$.

But it cannot be partitioned into, say, 1 independent set and 1 clique.

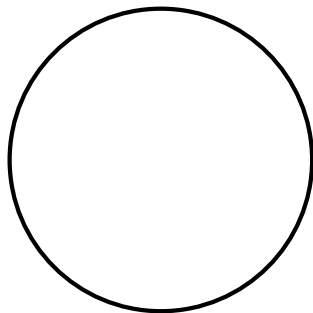
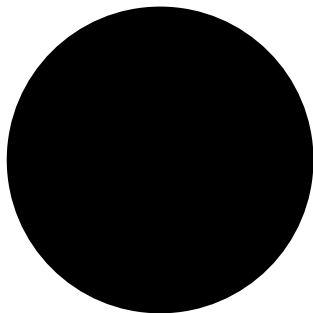


An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

So, $\chi_B(C_5) \leq 3$.

But it cannot be partitioned into, say, 1 independent set and 1 clique.



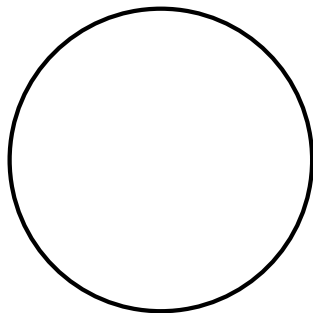
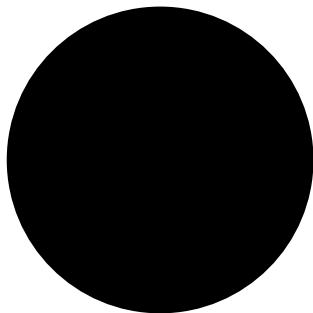
So, $\chi_B(C_5) \geq 3$.

An example: C_5

We will compute χ_B for the 5-cycle, C_5 .

So, $\chi_B(C_5) = 3$.

But it cannot be partitioned into, say, 1 independent set and 1 clique.



So, $\chi_B(C_5) = 3$.

Since $\chi_B(C_5) = 3$, and C_5 is self-complementary, the theorem gives

$$d_{\text{Forb}(C_5)}^* = \frac{1}{2(\chi_B(H) - 1)}.$$

Since $\chi_B(C_5) = 3$, and C_5 is self-complementary, the theorem gives

$$d_{\text{Forb}(C_5)}^* = \frac{1}{2(3-1)}.$$

Since $\chi_B(C_5) = 3$, and C_5 is self-complementary, the theorem gives

$$d_{\text{Forb}(C_5)}^* = \frac{1}{4}.$$

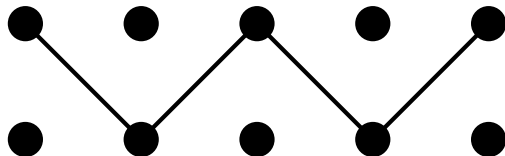
In fact,

Theorem (Marchant (2011))

$$\text{ed}_{\text{Forb}(C_5)}(p) = \min \left\{ \frac{p}{2}, \frac{1-p}{2} \right\}.$$

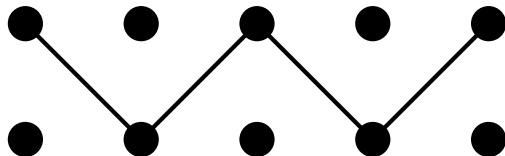
The biology problem

Recall the biology-inspired problem of removing induced copies of P_5 from a bipartite graph.



The biology problem

Recall the biology-inspired problem of removing induced copies of P_5 from a bipartite graph.



Theorem

Let H be a bipartite graph that is neither complete nor empty.
Let $G(n, n; 1/2)$ denote the random bipartite graph on an $n \times n$ vertex set with edge-probability $1/2$.

The number of edits required to ensure no induced copies of H is, a.a.s.,

$$\left(\frac{1}{2} - o(1)\right) n^2.$$

The biology problem

Theorem

Let H be a bipartite graph that is neither complete nor empty.

Let $G(n, n; 1/2)$ denote the random bipartite graph on an $n \times n$ vertex set with edge-probability $1/2$.

The number of edits required to ensure no induced copies of H is, a.a.s.,

$$\left(\frac{1}{2} - o(1)\right) n^2.$$

The result follows from the regularity lemma.

For the non-bipartite setting, one must use a strong version of the regularity lemma, such as Alon-Fischer-Krivelevich-M. Szegedy (2000).

The biology problem

Theorem

Let H be a bipartite graph that is neither complete nor empty.

Let $G(n, n; 1/2)$ denote the random bipartite graph on an $n \times n$ vertex set with edge-probability $1/2$.

The number of edits required to ensure no induced copies of H is, a.a.s.,

$$\left(\frac{1}{2} - o(1)\right) n^2.$$

The result follows from the regularity lemma.

For the non-bipartite setting, one must use a strong version of the regularity lemma, such as Alon-Fischer-Krivelevich-M. Szegedy (2000).

Edit distance results are a consequence of Szemerédi's regularity lemma, quadratic programming, and structural properties of graphs.

Theorem

Let H be a graph with $\chi = \chi(H)$ and $\bar{\chi} = \chi(\bar{H})$ and let $\mathcal{H} = \text{Forb}(H)$.

- If $\chi \geq 2$, then $\text{ed}_{\mathcal{H}}(p) \leq \frac{p}{\chi-1}$.
- If $\bar{\chi} \geq 2$, then $\text{ed}_{\mathcal{H}}(p) \leq \frac{1-p}{\bar{\chi}-1}$.

Theorem

Let H be a graph with $\chi = \chi(H)$ and $\bar{\chi} = \chi(\bar{H})$ and let $\mathcal{H} = \text{Forb}(H)$.

- If $\chi \geq 2$, then $\text{ed}_{\mathcal{H}}(p) \leq \frac{p}{\chi-1}$.
- If $\bar{\chi} \geq 2$, then $\text{ed}_{\mathcal{H}}(p) \leq \frac{1-p}{\bar{\chi}-1}$.

Theorem

Let H be a graph and (r, s) be a pair of integers such that H **cannot** be partitioned into r independent sets and s cliques and let $\mathcal{H} = \text{Forb}(H)$.

Then,

$$\text{ed}_{\mathcal{H}}(p) \leq \frac{p(1-p)}{r(1-p) + sp}.$$

Edit distance functions for split graphs

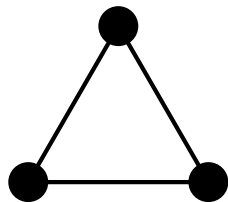
A **SPLIT GRAPH** is a graph that can be partitioned into one independent set and one clique.

Theorem

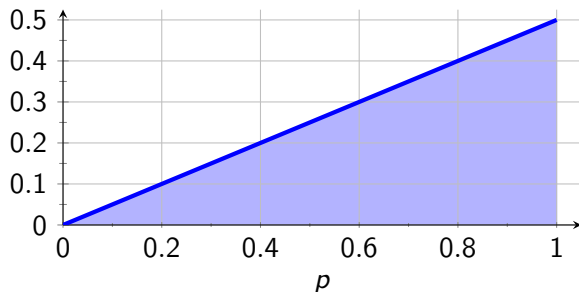
Let H be a split graph with $\alpha = \alpha(H)$, the independence number of H , and $\omega = \omega(H)$, the clique number of H , and let $\mathcal{H} = \text{Forb}(H)$.

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{\omega - 1}, \frac{1 - p}{\alpha - 1} \right\}; \quad p_{\mathcal{H}}^* = \frac{\omega - 1}{\alpha + \omega - 2}, \quad d_{\mathcal{H}}^* = \frac{1}{\alpha + \omega - 2}.$$

Edit distance functions for C_h



The cycle C_3 .



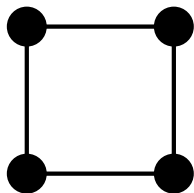
Theorem (Mantel (1907))

Let $h = 3$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. Then,

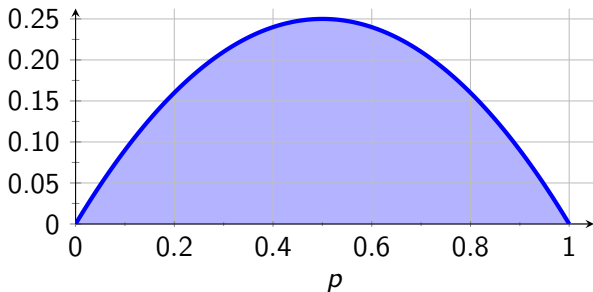
$$\text{ed}_{\mathcal{H}}(p) = \frac{p}{2};$$

$$p_{\mathcal{H}}^* = 1, \quad d_{\mathcal{H}}^* = \frac{1}{2}.$$

Edit distance functions for C_h



The cycle C_4 .

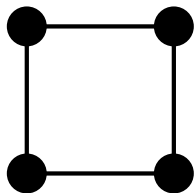


Theorem (Marchant-Thomason (2010))

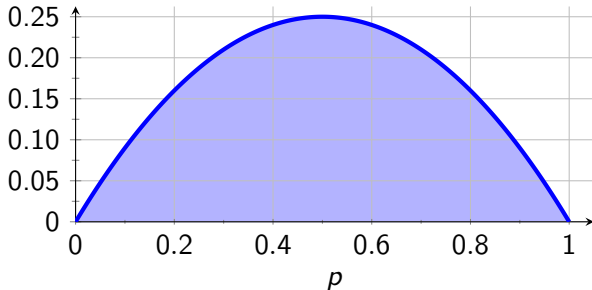
Let $h = 4$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. Then,

$$\text{ed}_{\mathcal{H}}(p) = p(1 - p); \quad p_{\mathcal{H}}^* = \frac{1}{2}, \quad d_{\mathcal{H}}^* = \frac{1}{4}.$$

Edit distance functions for C_h



The cycle C_4 .



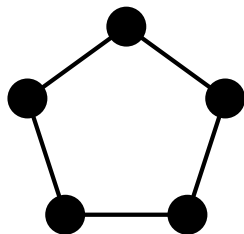
Theorem (Marchant-Thomason (2010))

Let $h = 4$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. Then,

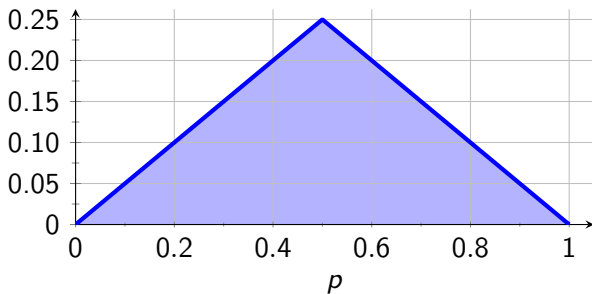
$$\text{ed}_{\mathcal{H}}(p) = p(1 - p); \quad p_{\mathcal{H}}^* = \frac{1}{2}, \quad d_{\mathcal{H}}^* = \frac{1}{4}.$$

The way editing is done is to split the vertex set into two different sized pieces, delete edges in one, and add edges in the other.

Edit distance functions for C_h



The cycle C_5 .

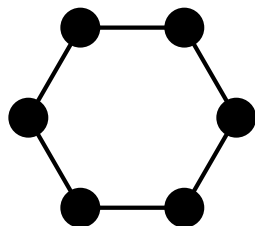


Theorem (Marchant (2011))

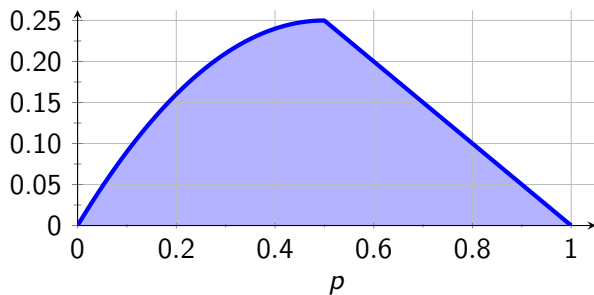
Let $h = 5$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. Then,

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{2}, \frac{1-p}{2} \right\}; \quad p_{\mathcal{H}}^* = \frac{1}{2}, \quad d_{\mathcal{H}}^* = \frac{1}{4}.$$

Edit distance functions for C_h



The cycle C_6 .

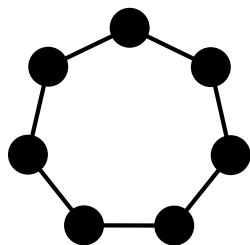


Theorem (M. (2013))

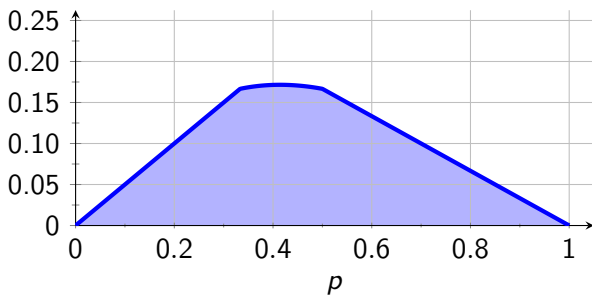
Let $h = 6$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. Then,

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ p(1-p), \frac{1-p}{2} \right\}; \quad p_{\mathcal{H}}^* = \frac{1}{2}, \quad d_{\mathcal{H}}^* = \frac{1}{4}.$$

Edit distance functions for C_h



The cycle C_7 .

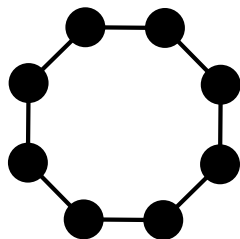


Theorem (Marchant (2011))

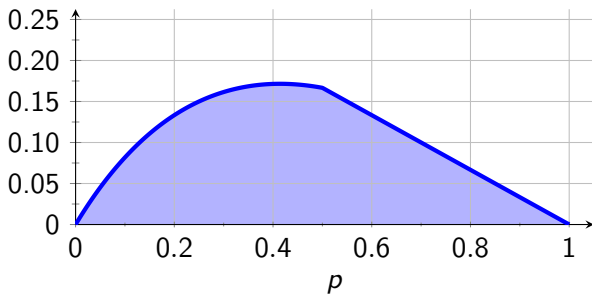
Let $h = 7$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. Then,

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{2}, \frac{p(1-p)}{1+p}, \frac{1-p}{3} \right\}; \quad p_{\mathcal{H}}^* = \sqrt{2} - 1, \quad d_{\mathcal{H}}^* = 3 - 2\sqrt{2}.$$

Edit distance functions for C_h



The cycle C_8 .

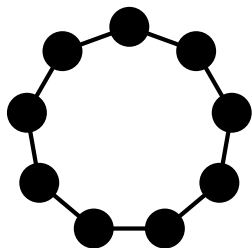


Theorem (M. (2013))

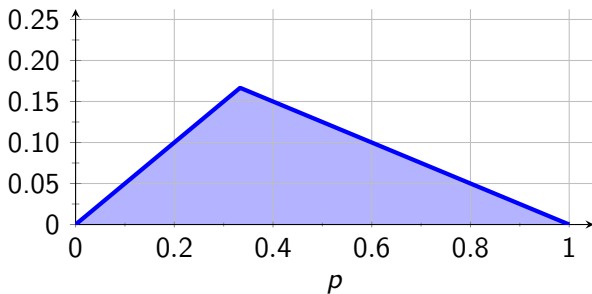
Let $h = 8$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. Then,

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p(1-p)}{1+p}, \frac{1-p}{3} \right\}; \quad p_{\mathcal{H}}^* = \sqrt{2} - 1, \quad d_{\mathcal{H}}^* = 3 - 2\sqrt{2}.$$

Edit distance functions for C_h



The cycle C_9 .

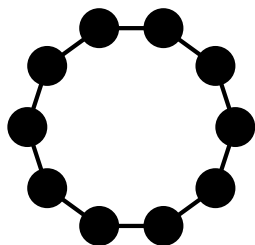


Theorem (M. (2013))

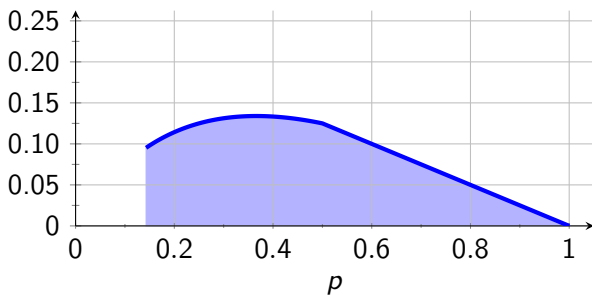
Let $h = 9$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. Then,

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{2}, \frac{1-p}{4} \right\}; \quad p_{\mathcal{H}}^* = \frac{1}{6}, \quad d_{\mathcal{H}}^* = \frac{1}{6}.$$

Edit distance functions for C_h



The cycle C_{10} .



Theorem (M. (2013))

Let $h = 10$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$. If $p \geq 1/7$, then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p(1-p)}{1+2p}, \frac{1-p}{4} \right\}; \quad p_{\mathcal{H}}^* = \frac{\sqrt{3}-1}{2}, \quad d_{\mathcal{H}}^* = \frac{2-\sqrt{3}}{2}.$$

Edit distance functions for C_h

Theorem (Peck (2013))

Let $h \geq 4$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$.

If h is odd and $p \in [0, 1]$, then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{2}, \frac{p(1-p)}{1 + (\lceil h/3 \rceil - 2)p}, \frac{1-p}{\lceil h/2 \rceil - 1} \right\}.$$

If h is even and $p \in [\lceil h/3 \rceil^{-1}, 1]$, then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p(1-p)}{1 + (\lceil h/3 \rceil - 2)p}, \frac{1-p}{\lceil h/2 \rceil - 1} \right\}.$$

Note: The function is known for all $p \in [0, 1]$ when $h = 6, 8$.
The function is known for $p \in [1/7, 1]$ when $h = 10$.
The small values of p are still open for even cycles.

Edit distance functions for C_h

Theorem (Peck (2013))

Let $h \geq 4$, $H = C_h$, and let $\mathcal{H} = \text{Forb}(H)$.

If $h \notin \{4, 7, 8, 10, 16\}$, then

$$p_{\mathcal{H}}^* = \frac{1}{\lceil h/2 \rceil - \lceil h/3 \rceil + 1} \quad d_{\mathcal{H}}^* = \frac{\lceil h/2 \rceil - \lceil h/3 \rceil}{(\lceil h/2 \rceil - 1)(\lceil h/2 \rceil - \lceil h/3 \rceil + 1)}.$$

If $h = 4$, then $p_{\mathcal{H}}^* = \frac{1}{2} \quad d_{\mathcal{H}}^* = \frac{1}{4}.$

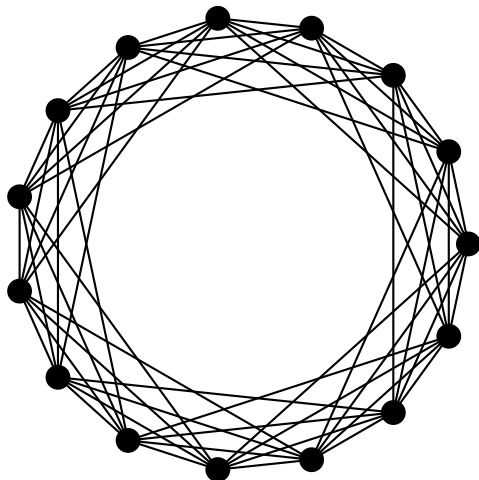
If $h \in \{7, 8\}$, then $p_{\mathcal{H}}^* = \sqrt{2} - 1 \quad d_{\mathcal{H}}^* = 3 - 2\sqrt{2}.$

If $h = 10$, then $p_{\mathcal{H}}^* = \frac{\sqrt{3} - 1}{2} \quad d_{\mathcal{H}}^* = \frac{2 - \sqrt{3}}{2}.$

If $h = 16$, then $p_{\mathcal{H}}^* = \frac{\sqrt{5} - 1}{4} \quad d_{\mathcal{H}}^* = \frac{2 - \sqrt{3}}{2}.$

Powers of cycles

Let C_h^t denote the t^{th} power of a cycle of length h , $h \geq 2t + 1$.



The graph C_{15}^4 .

Powers of cycles

Let C_h^t denote the t^{th} power of a cycle of length h , $h \geq 2t + 1$.

We can compute $\text{ed}_{\mathcal{H}}(p)$ for $\mathcal{H} = \text{Forb}(C_h^t)$ where $h \geq 2t^2 + 2t + 1$ but the expression is somewhat complex. A corollary is:

Theorem (Berikkyzy-M.-Peck (2018+))

Let $t \geq 1$, $h \geq 4t^2 + 10t + 24$, $H = C_h^t$, and let $\mathcal{H} = \text{Forb}(H)$.

Let $\ell_0 = \left\lceil \frac{h}{t+1} \right\rceil$, $\ell_t = \left\lceil \frac{h}{2t+1} \right\rceil$, and $p_t = \ell_t^{-1}$. Then

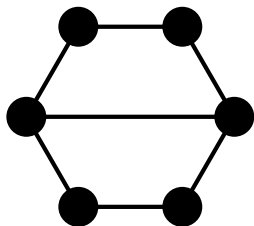
If $(t+1) \nmid h$ and $p \in [0, 1]$, then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{t+1}, \frac{p(1-p)}{t(1-p) + (\ell_t - 1)p}, \frac{1-p}{\ell_0 - 1} \right\}.$$

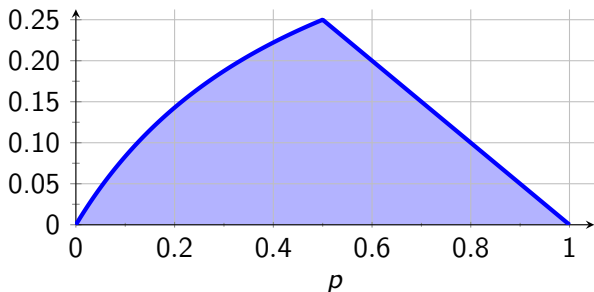
If $(t+1) \mid h$ and $p \in [p_t^{-1}, 1]$, then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p(1-p)}{t(1-p) + (\ell_t - 1)p}, \frac{1-p}{\ell_0 - 1} \right\}.$$

Interesting graphs



The graph C_6^* .



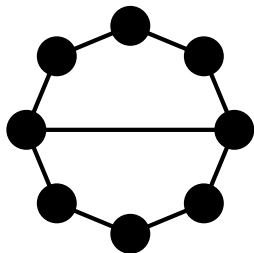
Theorem (Marchant-Thomason (2010))

Let $H = C_6^*$ be the 6-cycle with a diagonal and let $\mathcal{H} = \text{Forb}(H)$. Then

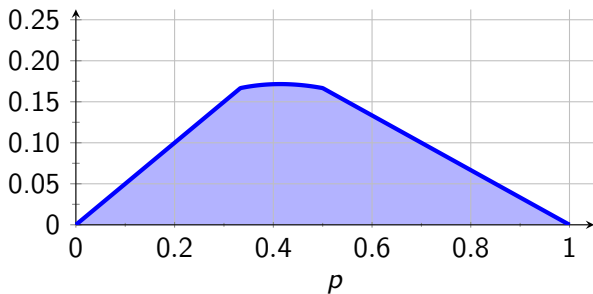
$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{1+2p}, \frac{1-p}{2} \right\}; \quad p_{\mathcal{H}}^* = \frac{1}{2}, \quad d_{\mathcal{H}}^* = \frac{1}{4}.$$

The editing scheme for $p < 1/2$ is more complicated than partitioning the vertex set and either deleting or adding inside the parts.

Interesting graphs



The graph C_8^* .

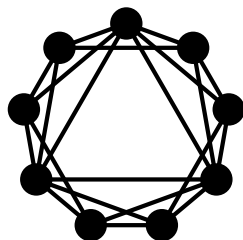


Theorem (Hu-Shi-Wei (2018+))

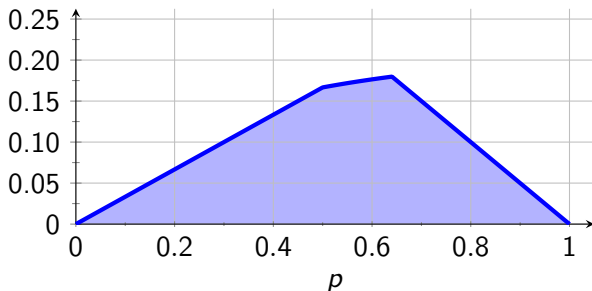
Let $H = C_8^*$ be the 8-cycle with a diagonal and let $\mathcal{H} = \text{Forb}(H)$. Then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{2}, \frac{p(1-p)}{1+p}, \frac{1-p}{3} \right\}; \quad p_{\mathcal{H}}^* = \sqrt{2} - 1, \quad d_{\mathcal{H}}^* = 3 - 2\sqrt{2}.$$

Interesting graphs



The graph H_9 .

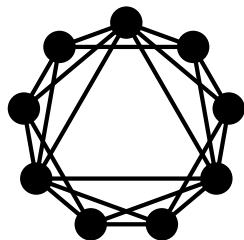


Theorem (M. (2015))

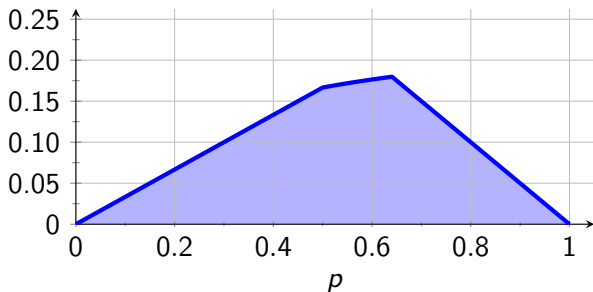
Let $H = H_9$ be C_9^2 plus a triangle and let $\mathcal{H} = \text{Forb}(H)$. Then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{3}, \frac{p}{1+4p}, \frac{1-p}{2} \right\}; \quad p_{\mathcal{H}}^* = \frac{1 + \sqrt{17}}{8}, \quad d_{\mathcal{H}}^* = \frac{7 - \sqrt{17}}{8}.$$

Interesting graphs



The graph H_9 .

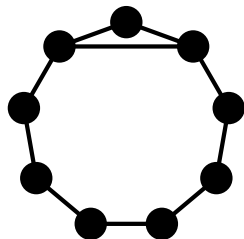


Theorem (M. (2015))

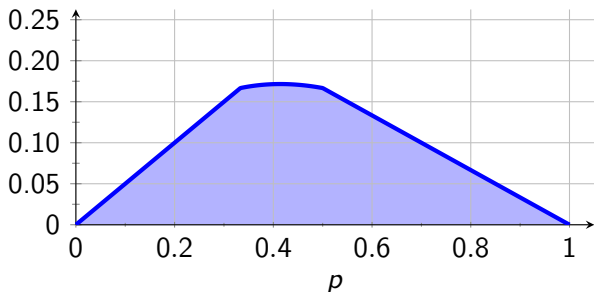
$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{3}, \frac{p}{1+4p}, \frac{1-p}{2} \right\}; \quad p_{\mathcal{H}}^* = \frac{1 + \sqrt{17}}{8}, \quad d_{\mathcal{H}}^* = \frac{7 - \sqrt{17}}{8}.$$

One of the editing schemes that results in $d_{\mathcal{H}}^*$ is a complex one. This was the first example of a complex scheme effecting $d_{\mathcal{H}}^*$.

Interesting graphs



The graph \tilde{C}_9 .

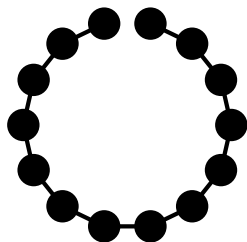


Theorem (Hu-Shi-Wei (2018+))

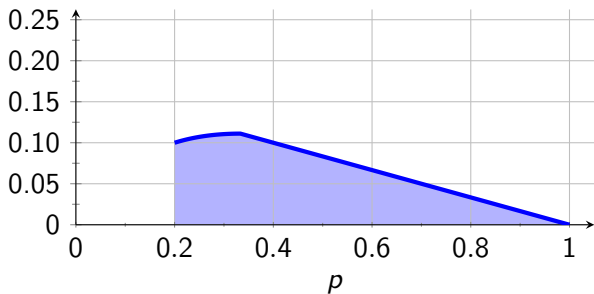
Let $H = \tilde{C}_h$ be the h -cycle with a single chord between a pair of vertices of distance two and let $\mathcal{H} = \text{Forb}(H)$. If $h \geq 9$, then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{2}, \frac{p(1-p)}{1 + (\lceil \frac{h-1}{3} \rceil - 2)p}, \frac{1-p}{\lceil \frac{h-1}{2} \rceil - 1} \right\}.$$

Interesting graphs



The path P_{14} .

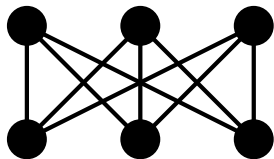


Theorem (Hu-Shi-Wei (2018+))

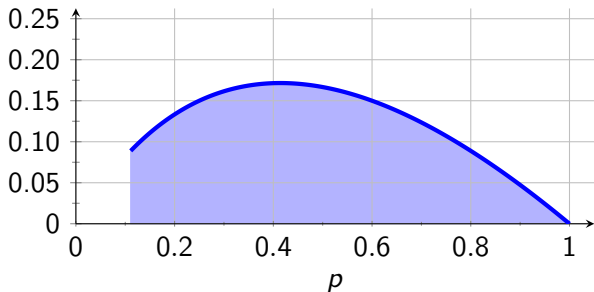
Let $H = P_h$ be the path on h vertices and let $\mathcal{H} = \text{Forb}(H)$. If $p \geq \lceil \frac{h-1}{3} \rceil^{-1}$, then

$$\text{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p(1-p)}{1 + (\lceil \frac{h-1}{3} \rceil - 2)p}, \frac{1-p}{\lceil \frac{h}{2} \rceil - 1} \right\}.$$

Interesting graphs



The graph $K_{3,3}$.



Theorem (Marchant-Thomason (2010))

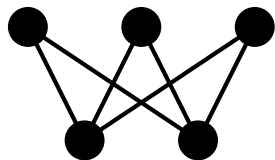
Let $H = K_{3,3}$ and let $\mathcal{H} = \text{Forb}(H)$. If $p \geq 1/9$, then

$$\text{ed}_{\mathcal{H}}(p) = \frac{p(1-p)}{1+p}; \quad p_{\mathcal{H}}^* = \sqrt{2} - 1, \quad d_{\mathcal{H}}^* = 3 - 2\sqrt{2}.$$

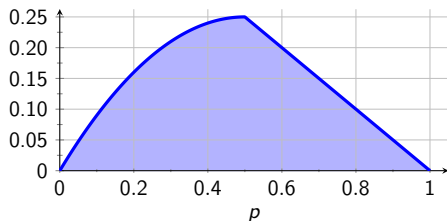
Interestingly, $\text{ed}_{\mathcal{H}}(p) < \frac{p(1-p)}{1+p}$ for $0 < p < \frac{1}{124}$.

By concavity, $\text{ed}_{\mathcal{H}}(p) \geq \frac{4}{5}p$ for all $0 \leq p \leq \frac{1}{9}$.

Edit distance functions for $K_{2,t}$



The graph $K_{2,3}$.

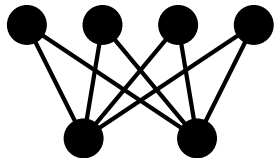


Theorem (M.-McKay (2014))

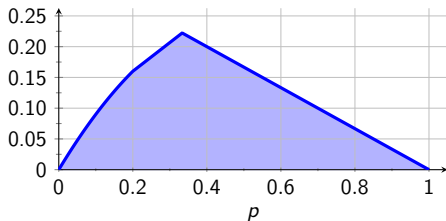
Let $t \geq 3$, $H = K_{2,t}$ and let $\mathcal{H} = \text{Forb}(H)$.

- If $t = 3$, then $\text{ed}_{\mathcal{H}}(p) = \min \left\{ p(1-p), \frac{1-p}{2} \right\}$.

Edit distance functions for $K_{2,t}$



The graph $K_{2,4}$.

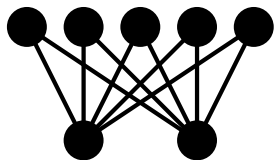


Theorem (M.-McKay (2014))

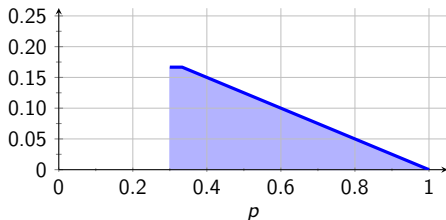
Let $t \geq 3$, $H = K_{2,t}$ and let $\mathcal{H} = \text{Forb}(H)$.

- If $t = 3$, then $\text{ed}_{\mathcal{H}}(p) = \min \left\{ p(1-p), \frac{1-p}{2} \right\}$.
- If $t = 4$, then $\text{ed}_{\mathcal{H}}(p) = \min \left\{ p(1-p), \frac{7p+1}{15}, \frac{1-p}{3} \right\}$.

Edit distance functions for $K_{2,t}$



The graph $K_{2,5}$.

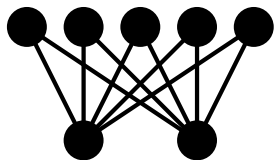


Theorem (M.-McKay (2014))

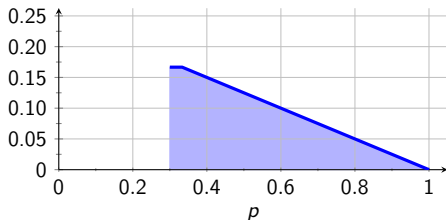
Let $t \geq 3$, $H = K_{2,t}$ and let $\mathcal{H} = \text{Forb}(H)$.

- If $t = 3$, then $\text{ed}_{\mathcal{H}}(p) = \min \left\{ p(1-p), \frac{1-p}{2} \right\}$.
- If $t = 4$, then $\text{ed}_{\mathcal{H}}(p) = \min \left\{ p(1-p), \frac{7p+1}{15}, \frac{1-p}{3} \right\}$.
- If $t \geq 5$ and is odd, then $d_{\mathcal{H}}^* = \frac{1}{t+1}$ and $p_{\mathcal{H}}^* \supseteq \left[\frac{2t-1}{t(t+1)}, \frac{2}{t+1} \right]$.

Edit distance functions for $K_{2,t}$



The graph $K_{2,5}$.



Theorem (M.-McKay (2014))

Let $t \geq 3$, $H = K_{2,t}$ and let $\mathcal{H} = \text{Forb}(H)$.

- If $t = 3$, then $\text{ed}_{\mathcal{H}}(p) = \min \left\{ p(1-p), \frac{1-p}{2} \right\}$.
- If $t = 4$, then $\text{ed}_{\mathcal{H}}(p) = \min \left\{ p(1-p), \frac{7p+1}{15}, \frac{1-p}{3} \right\}$.
- If $t \geq 5$ and is odd, then $d_{\mathcal{H}}^* = \frac{1}{t+1}$ and $p_{\mathcal{H}}^* \supseteq \left[\frac{2t-1}{t(t+1)}, \frac{2}{t+1} \right]$.
- If $t \geq 9$, then there is a $p_0 = p_0(t) < 1/2$ such that $\text{ed}_{\mathcal{H}}(p) < p(1-p)$ for all $p < p_0$.

What's next?

Part II:

- General theory for edit distance.
- Necessary lemmas and tools.

What's next?

Part II:

- General theory for edit distance.
- Necessary lemmas and tools.

Part III:

- Step-by-step computation of a new edit distance function.
- Open problems and conjectures.