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.

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 $e(T_{n,p}) \approx \frac{p-1}{p}\binom{n}{2}$.

Stability

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

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

Stability

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

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

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

Stability

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

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

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

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

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 $V = \{1, 2, \ldots, n\}$.

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 $V = \{1, 2, \ldots, n\}$.
- The distance between graphs G', G'' is the normalized edit distance.

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

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 $V = \{1, 2, \ldots, n\}$.
- The distance between graphs G', G'' is the normalized edit distance.

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

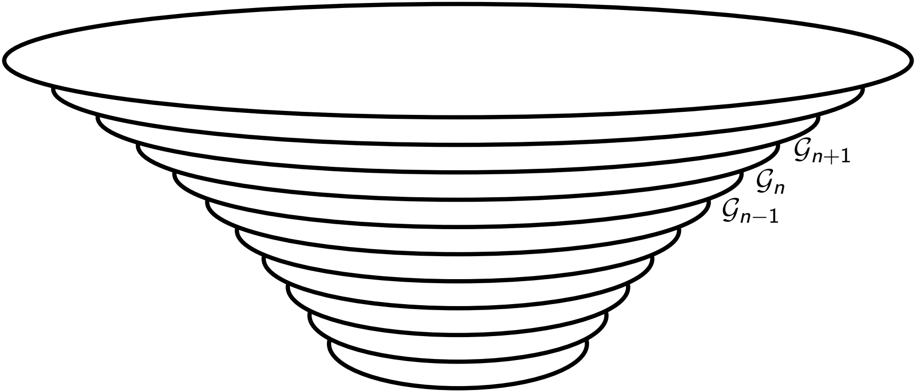
Metric Space

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

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

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:

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

Examples

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

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

Examples

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

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

Question

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

Examples

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

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

Question

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

Answer

$$0 \quad \text{or} \quad 2/5 \quad \text{or} \quad 3/5 \quad \text{or} \quad 1.$$

Examples

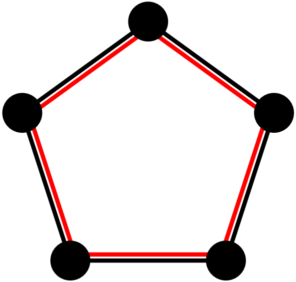
Question

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

Answer

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

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



Examples

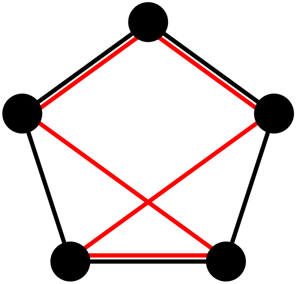
Question

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

Answer

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

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



Examples

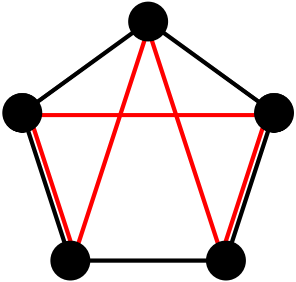
Question

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

Answer

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

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



Examples

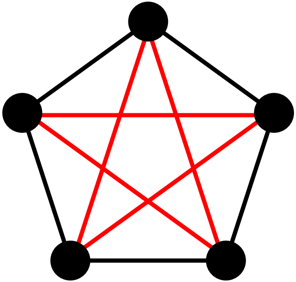
Question

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

Answer

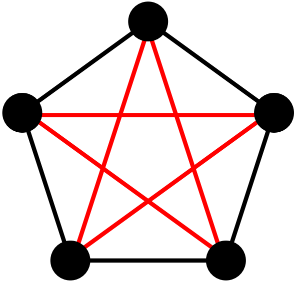
0 or 2/5 or 3/5 or *1*.

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.



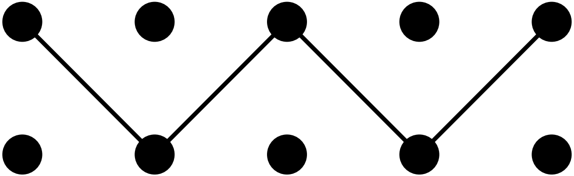
- We want the distance between graphs to be the distance between isomorphism classes.
- The distance between isomorphism classes of G and G' is the same as the distance between any copy of G and the isomorphism class of G'.
- So, we will consider the distance between a graph and a large set of graphs,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

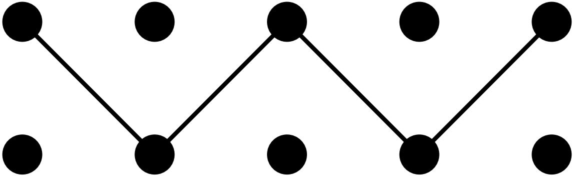


A question in biology

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?



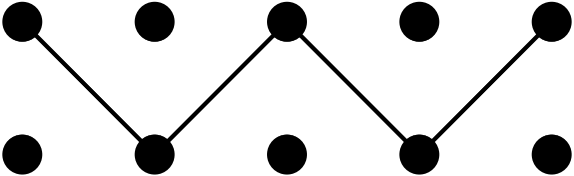
The motivation was from some work of Chen, Eulenstein, Fernández-Baca, and Sanderson on consensus trees in evolutionary biology.

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

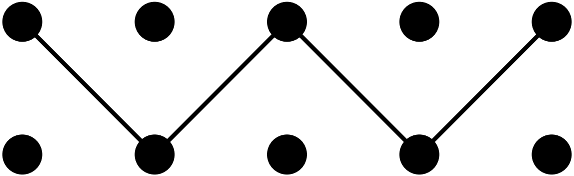
No.

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.

Editing

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

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

Editing

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

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

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

Editing

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

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

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

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

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

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., $G' = K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

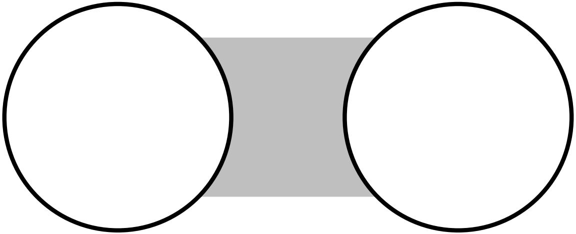
The graph achieving the maximum distance is the complete graph, K_n and

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

A classical problem

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

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



A classical problem

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

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

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

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

Cliques

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

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

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 $\mathcal{P}$

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

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 $\mathcal{P}$

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

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 $\mathcal{P}$

$$\begin{aligned}\mathrm{dist}(G,\mathcal{P}) &= \min\left\{\mathrm{dist}(G,G') : G' \in \mathcal{P}\right\}\\ &= \min\left\{|E(G)\triangle E(G')|/\tbinom{n}{2} : G' \in \mathcal{P}\right\}\end{aligned}$$

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 $\mathcal{P}$

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

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 $\mathcal{P}$

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

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 $\mathcal{H}$

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

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

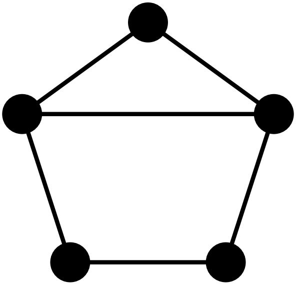
A HEREDITARY PROPERTY, $\mathcal{H}$, of graphs is one that holds under isomorphism and the deletion of vertices.

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

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 $\mathcal{H}$.

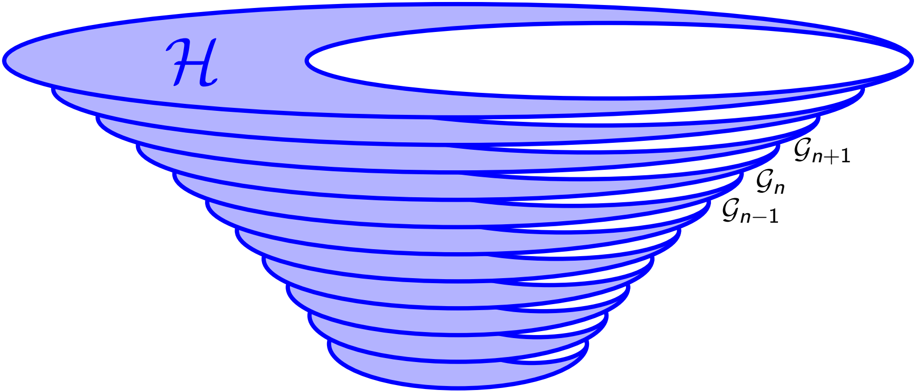


A 5-cycle as a subgraph, but no *induced* 5-cycle.

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.
- $\mathrm{Forb}(H)$

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

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

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

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:

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

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

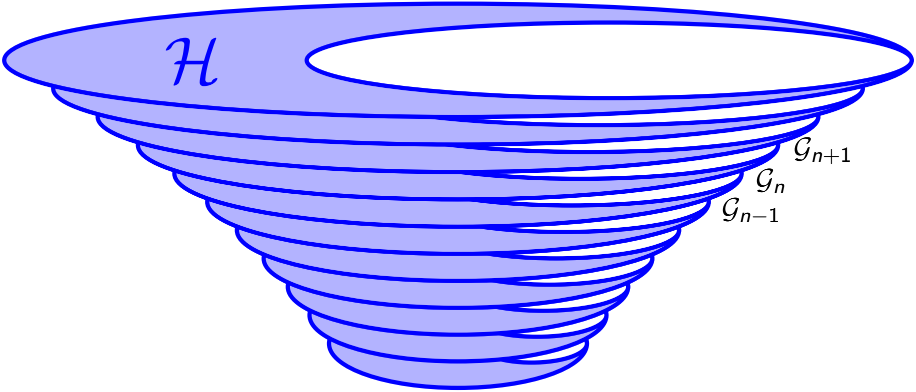
$$\mathcal{H} = \mathrm{Forb}(H), \text{ for some graph } H.$$

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

Understanding d^*

Problem

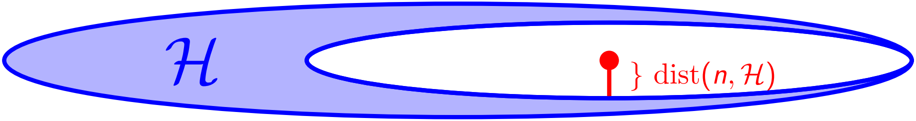
Compute $\mathrm{dist}(n, \mathcal{H})$ *for a hereditary property* $\mathcal{H}$.



Understanding d^*

Problem

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



Understanding d^*

Problem

Compute $\mathrm{dist}(n, \mathcal{H})$ *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 $\mathrm{dist}(n, \mathcal{H})$ *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*

$$d^*_{\mathcal{H}} = \lim_{n\to\infty} \mathrm{dist}(n, \mathcal{H}).$$

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

$$d^*_{\mathcal{H}} = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{dist}\left(G(n, p^*), \mathcal{H}\right)\right].$$

Understanding d^*

Theorem (Alon-Stav, 2008)

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

$$d^*_{\mathcal{H}} = \lim_{n\to\infty} \mathrm{dist}(n, \mathcal{H}).$$

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

$$d^*_{\mathcal{H}} = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{dist}\left(G(n,p^*),\mathcal{H}\right)\right].$$

Theorem (Concentration)

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

$$\Pr\left\{\left|\mathrm{dist}\left(G(n,p),\mathcal{H}\right) - \mu\right| \geq \epsilon\right\} \to 0.$$

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

Understanding d^*

Theorem (Alon-Stav, 2008)

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

$$d^*_{\mathcal{H}} = \lim_{n\to\infty} \mathrm{dist}(n, \mathcal{H}).$$

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

$$d^*_{\mathcal{H}} = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{dist}\left(G(n,p^*), \mathcal{H}\right)\right].$$

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

$$\mathrm{ed}_{\mathcal{H}}(p) = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{dist}\left(G(n,p),\mathcal{H}\right)\right]$$

then the limit exists and

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

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

$$\mathrm{ed}_{\mathcal{H}}(p) = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{dist}\left(G(n,p), \mathcal{H}\right)\right]$$

then the limit exists and

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

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

The Edit Distance Function

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

The Edit Distance Function

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

- Continuous and concave down.

The Edit Distance Function

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

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

The Edit Distance Function

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

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

The Edit Distance Function

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

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

The Edit Distance Function

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

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

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,

$$p^*_{\mathrm{Forb}(K_a+E_b)} = \frac{a-1}{a+b-1} \qquad d^*_{\mathrm{Forb}(K_a+E_b)} = \frac{1}{a+b-1}.$$

The Edit Distance Function

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

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

Theorem (Balogh-M., 2008)

$$p^*_{\mathrm{Forb}(K_{3,3})} = \sqrt{2} - 1 \qquad d^*_{\mathrm{Forb}(K_{3,3})} = 3 - 2\sqrt{2}.$$

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

Let $\mathcal{H} = \mathrm{Forb}(H)$ 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 $\mathrm{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \mathrm{Forb}(H)$ 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, $\chi_{\mathrm{B}}(H)$.

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

Let $\mathcal{H} = \mathrm{Forb}(H)$ 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, $\chi_{\mathrm{B}}(H)$.

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 $\mathrm{ed}_{\mathcal{H}}(1/2)$?

Let $\mathcal{H} = \mathrm{Forb}(H)$ 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, $\chi_{\mathrm{B}}(H)$.

Theorem (Prömel-Steger, 1993)

Let H be a graph.

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

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

Let $\mathcal{H} = \mathrm{Forb}(H)$ 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, $\chi_{\mathrm{B}}(H)$.

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

Let $\mathcal{H}$ be a hereditary property.

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

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

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

Let $\mathcal{H} = \mathrm{Forb}(H)$ 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, $\chi_{\mathrm{B}}(H)$.

Theorem (Axenovich-Kézdy-M., 2008)

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

$$\mathrm{ed}_{\mathcal{H}}(1/2) = \frac{1}{2(\chi_{\mathrm{B}}(H) - 1)} \leq d^*_{\mathcal{H}}.$$

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

Let $\mathcal{H} = \mathrm{Forb}(H)$ 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, $\chi_{\mathrm{B}}(H)$.

Theorem (Axenovich-Kézdy-M., 2008)

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

$$\mathrm{ed}_{\mathcal{H}}(1/2) = \frac{1}{2(\chi_{\mathrm{B}}(H) - 1)} \leq d^*_{\mathcal{H}}.$$

Moreover, equality holds if H is self-complementary.

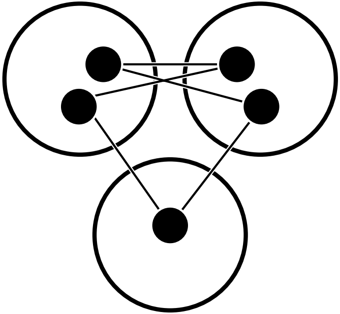
An example: C_5

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 $r + s = 3$:

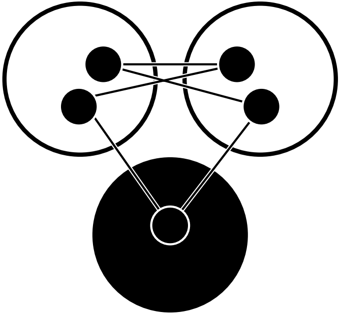


The 5-cycle can be partitioned into
3 independent sets and 0 cliques.

An example: C_5

We will compute χ_{B} for the 5-cycle, C_5.

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

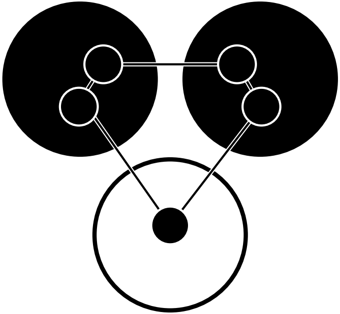


The 5-cycle can be partitioned into
2 independent sets and 1 cliques.

An example: C_5

We will compute χ_{B} for the 5-cycle, C_5.

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

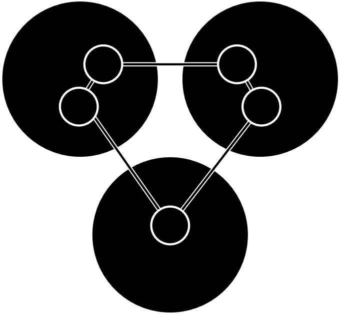


The 5-cycle can be partitioned into
1 independent sets and 2 cliques.

An example: C_5

We will compute χ_{B} for the 5-cycle, C_5.

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

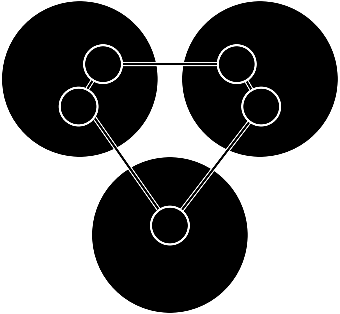


The 5-cycle can be partitioned into
0 independent sets and 3 cliques.

An example: C_5

We will compute χ_{B} for the 5-cycle, C_5.

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



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

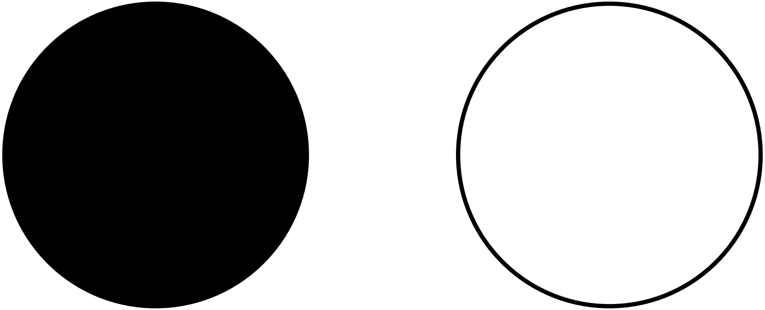
So, $\chi_{\mathrm{B}}(C_5) \leq 3$.

An example: C_5

We will compute χ_{B} for the 5-cycle, C_5.

So, $\chi_{\mathrm{B}}(C_5) \leq 3$.

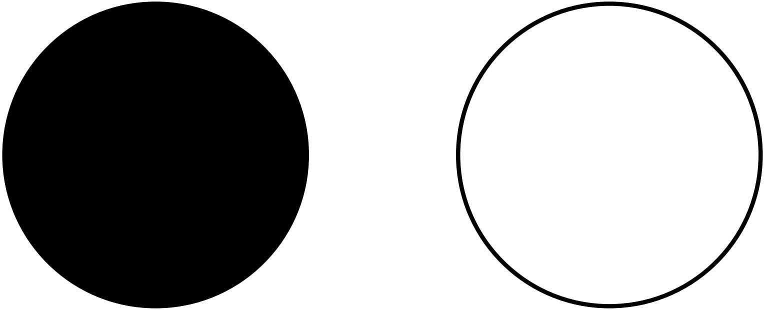
But it cannot be partitioned into, say, 1 independent set and 1 clique.



We will compute χ_{B} for the 5-cycle, C_5.

So, $\chi_{\mathrm{B}}(C_5) \leq 3$.

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



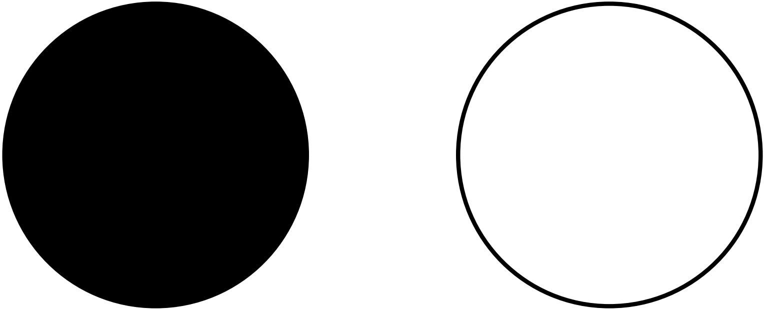
So, $\chi_{\mathrm{B}}(C_5) \geq 3$.

An example: C_5

We will compute χ_{B} for the 5-cycle, C_5.

So, $\chi_{\mathrm{B}}(C_5) = 3$.

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



So, $\chi_{\mathrm{B}}(C_5) = 3$.

Edit distance for C_5

Since $\chi_{\mathrm{B}}(C_5) = 3$, and C_5 is self-complementary, the theorem gives

$$d^*_{\mathrm{Forb}(C_5)} = \frac{1}{2(\chi_{\mathrm{B}}(H) - 1)}.$$

Edit distance for C_5

Since $\chi_{\mathrm{B}}(C_5) = 3$, and C_5 is self-complementary, the theorem gives

$$d^*_{\mathrm{Forb}(C_5)} = \frac{1}{2(3-1)}.$$

Edit distance for C_5

Since $\chi_{\mathrm{B}}(C_5) = 3$, and C_5 is self-complementary, the theorem gives

$$d^*_{\mathrm{Forb}(C_5)} = \frac{1}{4}.$$

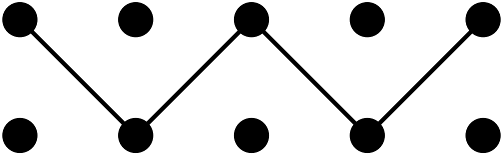
In fact,

Theorem (Marchant (2011))

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

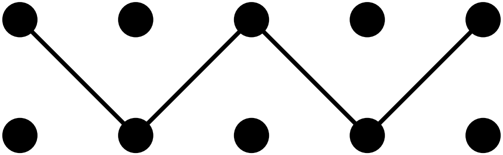
The biology problem

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.

Simple bounds

Theorem

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

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

Simple bounds

Theorem

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

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

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 $\mathcal{H} = \mathrm{Forb}(H)$. Then,

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

Edit distance functions for split graphs

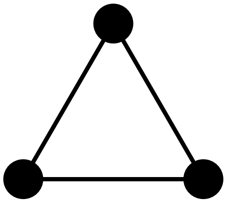
A SPLIT GRAPH is a graph that can be partitioned into one independent set and one clique.

Theorem

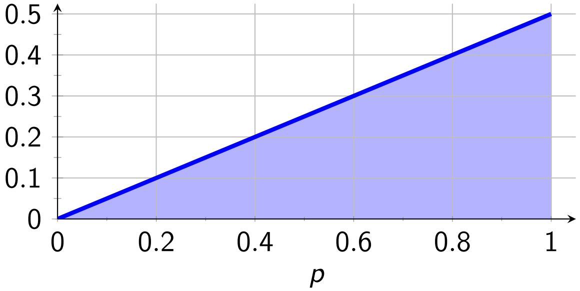
Let H be a split graph with $\alpha = \alpha(H)$, the independence number of H, and $\omega = \omega(H)$, the clique number of H, and let $\mathcal{H} = \mathrm{Forb}(H)$.

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

Edit distance functions for C_h



The cycle C_3.

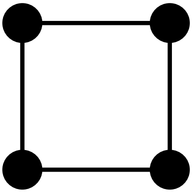


Theorem (Mantel (1907))

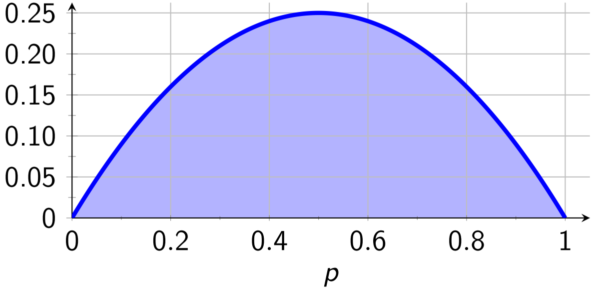
Let $h = 3$, $H = C_h$, and let $\mathcal{H} = \mathrm{Forb}(H)$. Then,

$$\mathrm{ed}_{\mathcal{H}}(p) = \frac{p}{2}; \qquad p^*_{\mathcal{H}} = 1, \quad d^*_{\mathcal{H}} = \frac{1}{2}.$$

Edit distance functions for C_h



The cycle C_4.

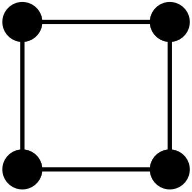


Theorem (Marchant-Thomason (2010))

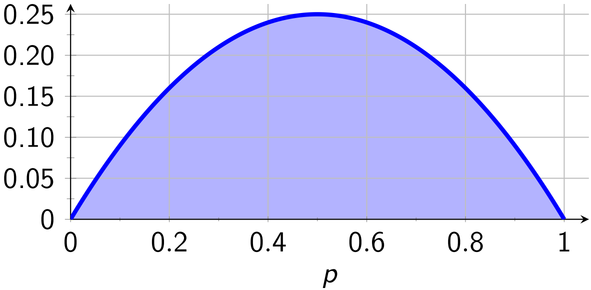
Let $h = 4$, $H = C_h$, *and let* $\mathcal{H} = \mathrm{Forb}(H)$. *Then,*

$$\mathrm{ed}_{\mathcal{H}}(p) = p(1-p); \qquad p^*_{\mathcal{H}} = \frac{1}{2}, \quad d^*_{\mathcal{H}} = \frac{1}{4}.$$

Edit distance functions for C_h



The cycle C_4.



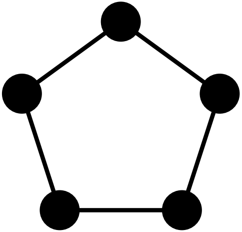
Theorem (Marchant-Thomason (2010))

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

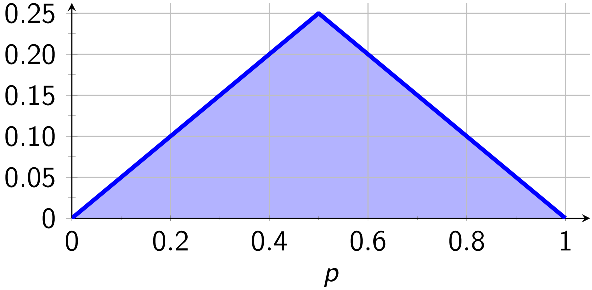
$$\mathrm{ed}_{\mathcal{H}}(p) = p(1-p); \qquad p^*_{\mathcal{H}} = \frac{1}{2}, \quad d^*_{\mathcal{H}} = \frac{1}{4}.$$

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.

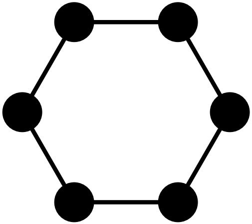


Theorem (Marchant (2011))

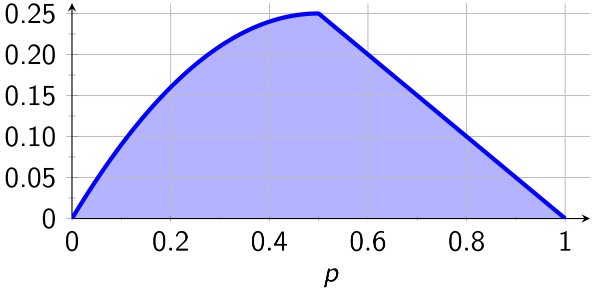
Let $h = 5$, $H = C_h$, and let $\mathcal{H} = \mathrm{Forb}(H)$. Then,

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

Edit distance functions for C_h



The cycle C_6.

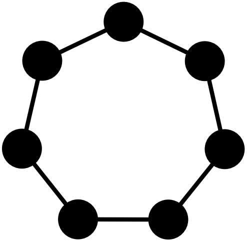


Theorem (M. (2013))

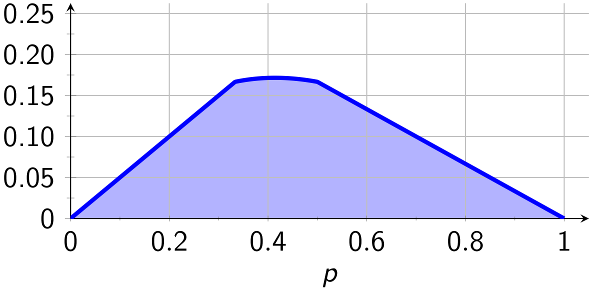
Let $h = 6$, $H = C_h$, and let $\mathcal{H} = \mathrm{Forb}(H)$. Then,

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

Edit distance functions for C_h



The cycle C_7.

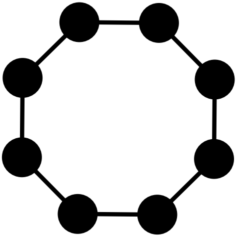


Theorem (Marchant (2011))

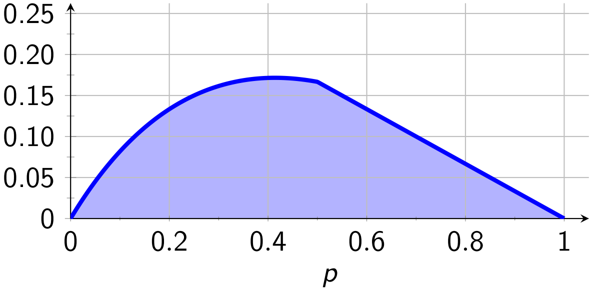
Let $h = 7$, $H = C_h$, *and let* $\mathcal{H} = \mathrm{Forb}(H)$. *Then,*

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

Edit distance functions for C_h



The cycle C_8.

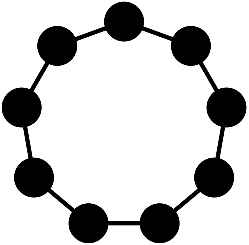


Theorem (M. (2013))

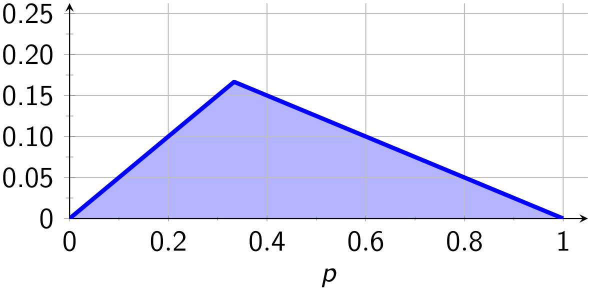
Let $h = 8$, $H = C_h$, and let $\mathcal{H} = \mathrm{Forb}(H)$. Then,

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

Edit distance functions for C_h



The cycle C_9.

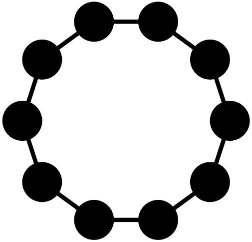


Theorem (M. (2013))

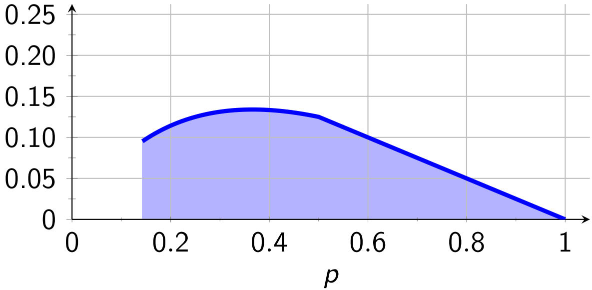
Let $h = 9$, $H = C_h$, *and let* $\mathcal{H} = \mathrm{Forb}(H)$. *Then,*

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

Edit distance functions for C_h



The cycle C_{10}.



Theorem (M. (2013))

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

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

Edit distance functions for C_h

Theorem (Peck (2013))

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

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

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

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

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

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* $\mathcal{H} = \mathrm{Forb}(H)$.
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* $p^*_{\mathcal{H}} = \frac{1}{2}$ $d^*_{\mathcal{H}} = \frac{1}{4}$.

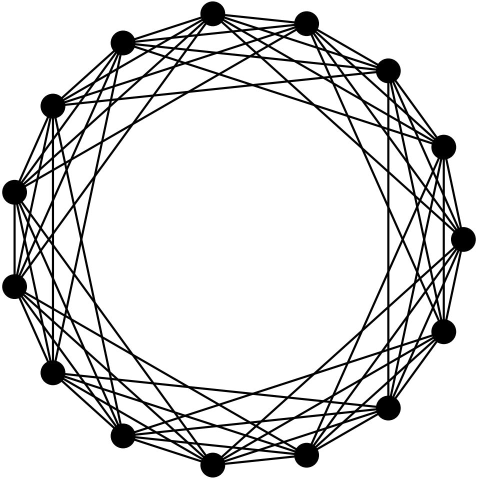
If $h \in \{7, 8\}$, *then* $p^*_{\mathcal{H}} = \sqrt{2} - 1$ $d^*_{\mathcal{H}} = 3 - 2\sqrt{2}$.

If $h = 10$, *then* $p^*_{\mathcal{H}} = \frac{\sqrt{3} - 1}{2}$ $d^*_{\mathcal{H}} = \frac{2 - \sqrt{3}}{2}$.

If $h = 16$, *then* $p^*_{\mathcal{H}} = \frac{\sqrt{5} - 1}{4}$ $d^*_{\mathcal{H}} = \frac{2 - \sqrt{3}}{2}$.

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

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

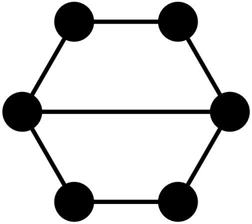
Let $t \geq 1$, $h \geq 4t^2+10t+24$, $H = C_h^t$, and let $\mathcal{H} = \mathrm{Forb}(H)$. Let $\ell_0 = \left\lceil \frac{h}{t+1} \right\rceil$, $\ell_t = \left\lceil \frac{h}{2t+1} \right\rceil$, and $p_t = \ell_t^{-1}$. Then

If $(t+1) \not| \, h$ and $p \in [0,1]$, then

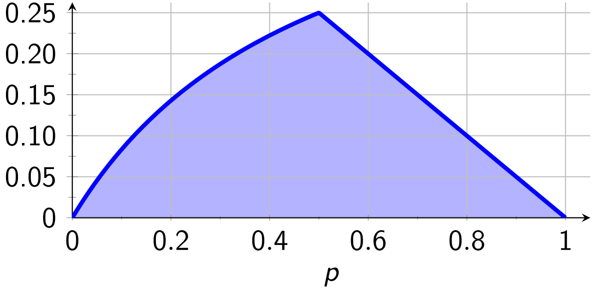
$$\mathrm{ed}_{\mathcal{H}}(p) = \min \left\{ \frac{p}{t+1}, \frac{p(1-p)}{t(1-p)+(\ell_t-1)\,p}, \frac{1-p}{\ell_0-1} \right\}.$$

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

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



The graph C_6^*.



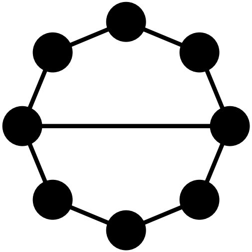
Theorem (Marchant-Thomason (2010))

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

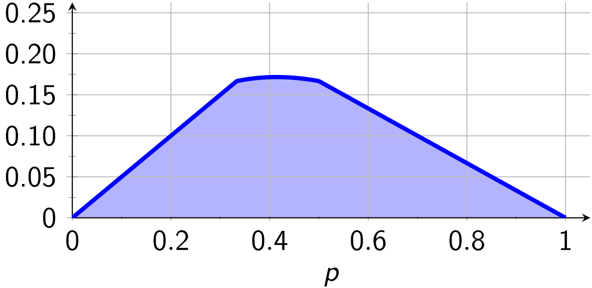
$$\mathrm{ed}_{\mathcal{H}}(p) = \min\left\{\frac{p}{1+2p}, \frac{1-p}{2}\right\}; \qquad p_{\mathcal{H}}^* = \frac{1}{2}, \quad d_{\mathcal{H}}^* = \frac{1}{4}.$$

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^*.

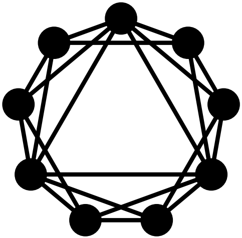


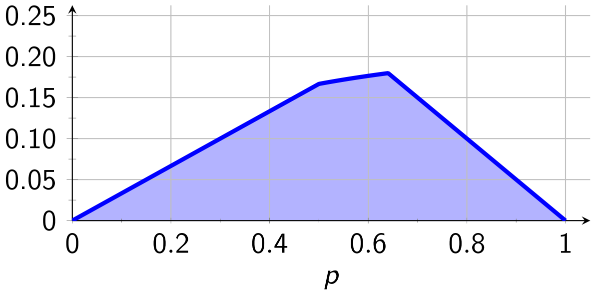
Theorem (Hu-Shi-Wei (2018+))

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

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

Interesting graphs


The graph H_9.

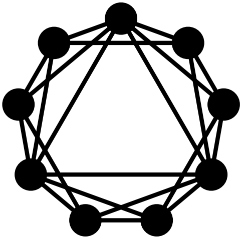


Theorem (M. (2015))

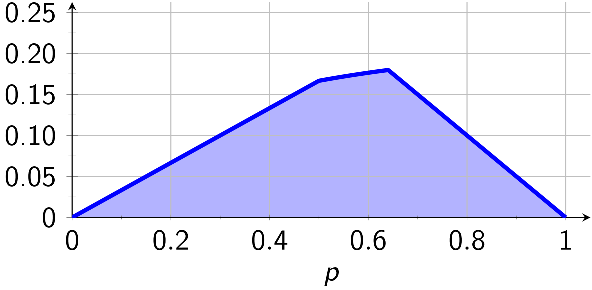
Let $H = H_9$ be C_9^2 plus a triangle and let $\mathcal{H} = \mathrm{Forb}(H)$. Then

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

Interesting graphs



The graph H_9.

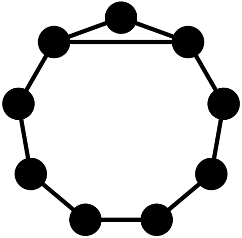


Theorem (M. (2015))

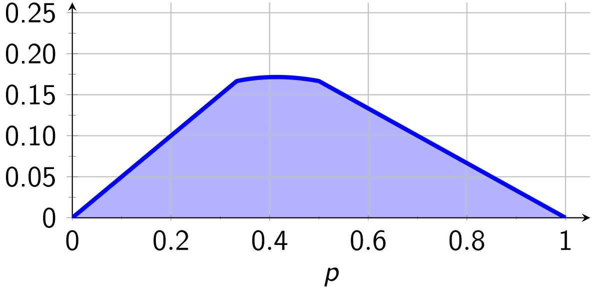
$$\mathrm{ed}_{\mathcal{H}}(p) = \min\left\{\frac{p}{3}, \frac{p}{1+4p}, \frac{1-p}{2}\right\}; \quad p^*_{\mathcal{H}} = \frac{1+\sqrt{17}}{8}, \quad d^*_{\mathcal{H}} = \frac{7-\sqrt{17}}{8}.$$

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 $\widetilde{C}_9$.

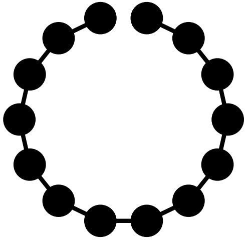


Theorem (Hu-Shi-Wei (2018+))

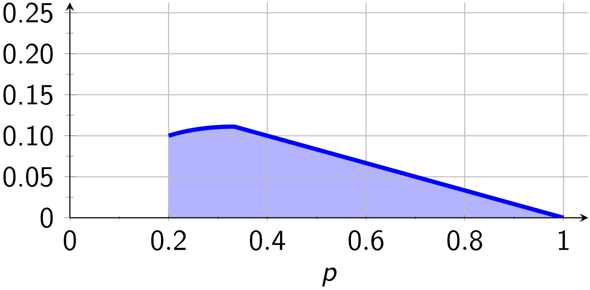
Let $H = \widetilde{C}_h$ be the h-cycle with a single chord between a pair of vertices of distance two and let $\mathcal{H} = \mathrm{Forb}(H)$. If $h \geq 9$, then

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

Interesting graphs



The path P_{14}.

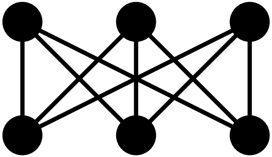


Theorem (Hu-Shi-Wei (2018+))

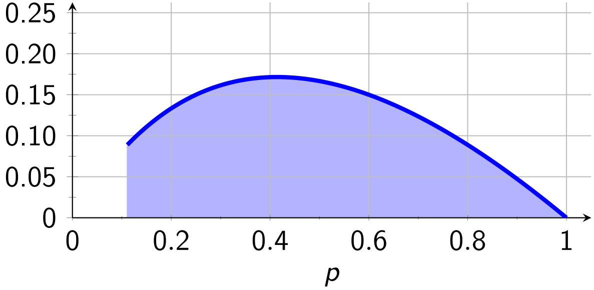
Let $H = P_h$ be the path on h vertices and let $\mathcal{H} = \mathrm{Forb}(H)$. If $p \geq \left\lceil \frac{h-1}{3} \right\rceil^{-1}$, then

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

Interesting graphs



The graph $K_{3,3}$.



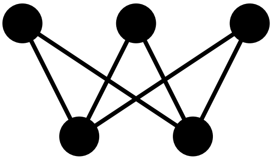
Theorem (Marchant-Thomason (2010))

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

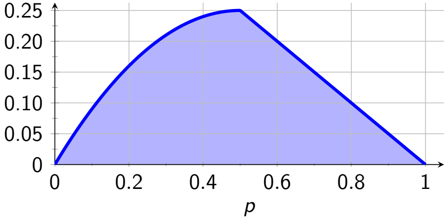
$$\mathrm{ed}_{\mathcal{H}}(p) = \frac{p(1-p)}{1+p}; \qquad p^*_{\mathcal{H}} = \sqrt{2} - 1, \quad d^*_{\mathcal{H}} = 3 - 2\sqrt{2}.$$

Interestingly, $\mathrm{ed}_{\mathcal{H}}(p) < \frac{p(1-p)}{1+p}$ for $0 < p < \frac{1}{124}$.
By concavity, $\mathrm{ed}_{\mathcal{H}}(p) \geq \frac{4}{5}p$ for all $0 \leq p \leq \frac{1}{9}$.

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



The graph $K_{2,3}$.

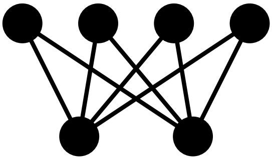


Theorem (M.-McKay (2014))

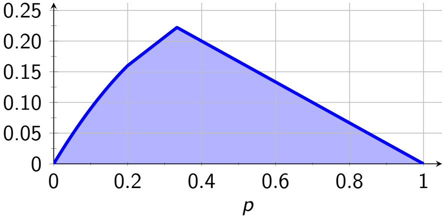
Let $t \geq 3$, $H = K_{2,t}$ and let $\mathcal{H} = \mathrm{Forb}(H)$.

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

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



The graph $K_{2,4}$.

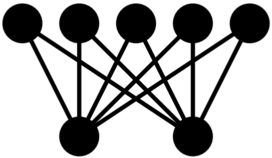


Theorem (M.-McKay (2014))

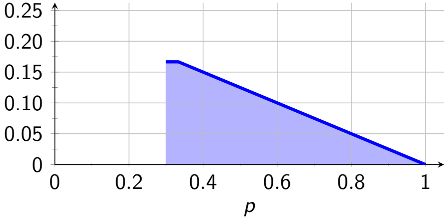
Let $t \geq 3$, $H = K_{2,t}$ *and let* $\mathcal{H} = \mathrm{Forb}(H)$.

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

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



The graph $K_{2,5}$.

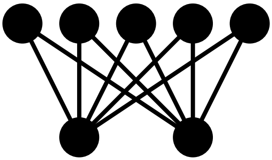


Theorem (M.-McKay (2014))

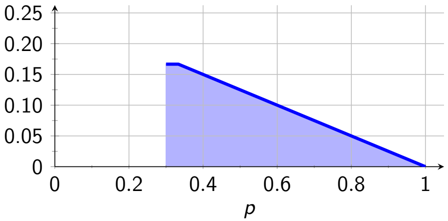
Let $t \geq 3$, $H = K_{2,t}$ and let $\mathcal{H} = \mathrm{Forb}(H)$.

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

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