Recent results on the edit distance of graphs

Ryan Martin

Iowa State University

Discrete Mathematics Seminar
University of Nebraska-Lincoln
Joint Work

This talk is based on joint work with:

Tracy McKay
ISU

József Balogh
Illinois

Maria Axenovich
ISU

André Kezdy
Louisville

Ryan Martin (Iowa State U.)
The edit distance of graphs

31 August 2010 2 / 23
Another view of a classical problem

Suppose we have an $n$-vertex graph $G$. 

Theorem (Mantel, 1907)

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

This bound is only achieved if $G'$ is complete bipartite.
Another view of a classical problem

Suppose we have an \( n \)-vertex graph \( G \).

We want to compute the fewest number of edge-additions plus edge-deletions to apply to \( G \) so that the resulting graph \( G' \) 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.
Another view of a classical problem

Suppose we have an $n$-vertex graph $G$.

We want to compute the fewest number of edge-additions plus edge-deletions to apply to $G$ so that the resulting graph $G'$ 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 $\left\lfloor \frac{n^2}{4} \right\rfloor$. This bound is only achieved if $G'$ is complete bipartite.*
Another view of a classical problem

Suppose we have an $n$-vertex graph $G$.

We want to compute the fewest number of edge-additions plus edge-deletions to apply to $G$ so that the resulting graph $G'$ 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 $\left\lfloor n^2/4 \right\rfloor$. This bound is only achieved if $G'$ is complete bipartite.*
So, if $G$ were the complete graph, it would require at least

$$\binom{n}{2} - \lfloor n^2/4 \rfloor$$

edge-deletions.
So, if $G$ were the complete graph, it would require at least

$$\binom{n}{2} - \lfloor n^2 / 4 \rfloor = \left(\left\lfloor \frac{n}{2} \right\rfloor \right)^2 + \left(\left\lceil \frac{n}{2} \right\rceil \right)^2$$

edge-deletions.
So, if $G$ were the complete graph, it would require at least

\[
\binom{n}{2} - \lfloor n^2 / 4 \rfloor = \left( \lfloor n/2 \rfloor \right)^2 + \left( \lceil n/2 \rceil \right)^2
\]

edge-deletions.

But for any graph $G$, we can delete at most this many edges and remain triangle-free:
So, if \( G \) were the complete graph, it would require at least

\[
\binom{n}{2} - \left\lfloor \frac{n^2}{4} \right\rfloor = \binom{\left\lfloor \frac{n}{2} \right\rfloor}{2} + \binom{\left\lceil \frac{n}{2} \right\rceil}{2}
\]

edge-deletions.

But for any graph \( G \), we can delete at most this many edges and remain triangle-free:

Partition the vertices in half and delete edges inside each part.
Results on triangles

So, the maximum number of changes required to remove triangles from $n$-vertex graph $G$ is

$$\left(\left\lfloor \frac{n}{2} \right\rfloor \right)^2 + \left(\left\lceil \frac{n}{2} \right\rceil \right)^2.$$

This is achieved by $G = K_n$. 

Results on triangles

So, the maximum number of changes required to remove triangles from $n$-vertex graph $G$ is

$$\binom{\lfloor n/2 \rfloor}{2} + \binom{\lceil n/2 \rceil}{2} \sim \frac{1}{2} \left( \frac{n^2}{2} \right).$$

This is achieved by $G = K_n$. 

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

Theorem (Turán, 1941)

If an $n$-vertex graph $G'$ has no copy of $K_\ell + 1$, then the number of edges $G'$ has is at most

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

This bound is only achieved if $G'$ is complete $\ell$-partite and $\ell | n$. 

So, the maximum number of changes required to remove copies of $K_\ell + 1$ from an $n$-vertex graph $G$ is

$$\sim n^{2/\ell} \sim \frac{1}{\ell} \left( \frac{n^2}{2} \right).$$

This is achieved by $G = K_n$. 

Ryan Martin (Iowa State U.)
Results on triangles

So, the maximum number of changes required to remove triangles from an $n$-vertex graph $G$ is

$$\left(\left\lfloor \frac{n}{2} \right\rfloor \right) + \left(\left\lceil \frac{n}{2} \right\rceil \right) \sim \frac{1}{2} \left( \frac{n^2}{2} \right).$$

This is achieved by $G = K_n$.

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

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

**Theorem (Turán, 1941)**

*If an $n$-vertex graph $G'$ has no copy of $K_{\ell+1}$, then the number of edges $G'$ has is at most*

$$\left(1 - \frac{1}{\ell}\right) \frac{n^2}{2}.$$ *This bound is only achieved if $G'$ is complete $\ell$-partite and $\ell \mid n$.***
Results on triangles

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

**Theorem (Turán, 1941)**

*If an $n$-vertex graph $G'$ has no copy of $K_{\ell+1}$, then the number of edges $G'$ has is at most*

\[
\left(1 - \frac{1}{\ell}\right) \frac{n^2}{2}.
\]

*This bound is only achieved if $G'$ is complete $\ell$-partite and $\ell \mid n$.

So, the maximum number of changes required to remove copies of $K_{\ell+1}$ from an $n$-vertex graph $G$ is

\[
\sim \frac{n^2}{2\ell}
\]
Results on triangles

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

**Theorem (Turán, 1941)**

*If an $n$-vertex graph $G'$ has no copy of $K_{\ell+1}$, then the number of edges $G'$ has is at most*

\[
\left(1 - \frac{1}{\ell}\right) \frac{n^2}{2}.
\]

*This bound is only achieved if $G'$ is complete $\ell$-partite and $\ell \mid n$.*

So, the maximum number of changes required to remove copies of $K_{\ell+1}$ from an $n$-vertex graph $G$ is

\[
\sim \frac{n^2}{2\ell} \sim \frac{1}{\ell} \binom{n}{2}.
\]

This is achieved by $G = K_n$. 

Ryan Martin (Iowa State U.)
Given: Labeled graphs $G$ and $G'$, each on $n$ vertices.
**Given:** Labeled graphs $G$ and $G'$, each on $n$ vertices.

**Definition**

The **EDIT DISTANCE BETWEEN $G$ AND $G'$**

$$\text{Dist}(G, G') = |E(G) \triangle E(G')|$$

is the size of the symmetric difference of the edge sets.
**Given:** Labeled graphs $G$ and $G'$, each on $n$ vertices.

**Definition**

The **EDIT DISTANCE BETWEEN $G$ AND $G'$**

$$\text{Dist}(G, G') = \left| E(G) \triangle E(G') \right|$$

is the size of the symmetric difference of the edge sets.

That is, it is the minimum number of edge-additions plus edge-deletions to transform $G$ into $G'$. 
Given: A labeled graph $G$ and a graph property $\mathcal{P}$.
A **GRAPH PROPERTY** is a set of graphs.
**Given:** A labeled graph $G$ and a graph property $\mathcal{P}$. A **GRAPH PROPERTY** is a set of graphs.

**Definition**

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

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

is the least edit distance of $G$ to a graph in $\mathcal{P}$.
Given: A labeled graph $G$ and a 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 number of edge-additions plus edge-deletions to transform $G$ into a member of $\mathcal{P}$. 
Given: A natural number $n$ and a graph property $\mathcal{P}$.
Extremal Edit Distance

**Given:** A natural number \( n \) and a graph property \( P \).

**Definition**

The **EDIT DISTANCE FROM \( P \)**

\[
\text{Dist}(n, P) = \max \{ \text{Dist}(G, P) : |V(G)| = n \}
\]

is the maximum edit distance of an \( n \)-vertex graph to a graph in \( P \).
**Given:** A natural number $n$ and a graph property $\mathcal{P}$.

**Definition**

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

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

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

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

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

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

- **Planar graphs.**
- **Line graphs of bipartite graphs.**
- **Chordal graphs:** graphs with no chordless cycle of length \( \geq 4 \).
- **Perfect graphs:** \( \chi = \omega \) for all induced subgraphs.
- **Forb** \((H)\): graphs with no induced copy of \( H \).
  - (Forb \((H)\) is a principal hereditary property.)

For the rest of this talk, all of our hereditary properties are principal; i.e., \( H = \text{Forb}(H) \), for some graph \( H \).
Hereditary Properties

**Definition**

A HEREDITARY PROPERTY, $\mathcal{H}$, of graphs is one that holds under 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 $\geq 4$.
- Perfect graphs: $\chi = \omega$ for all induced subgraphs.
- Forb($H$): graphs with no induced copy of $H$.

(Forb($H$) is a principal hereditary property.)

For the rest of this talk, all of our hereditary properties are principal; i.e., $H = \text{Forb}(H)$, for some graph $H$. 

Ryan Martin (Iowa State U.)
A hereditary property, $\mathcal{H}$, of graphs is one that holds under 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 not an induced graph.

A 5-cycle as a subgraph, but not an induced graph.
Hereditary Properties

Definition

A **hereditary property**, $\mathcal{H}$, of graphs is one that holds under 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 $\geq 4$.
- Perfect graphs: $\chi = \omega$ for all induced subgraphs.
- $\text{Forb}(\mathcal{H})$: graphs with no induced copy of $\mathcal{H}$.

($\text{Forb}(\mathcal{H})$ is a principal hereditary property.)

For the rest of this talk, all of our hereditary properties are principal; i.e., $\mathcal{H} = \text{Forb}(\mathcal{H})$, for some graph $\mathcal{H}$. 

Ryan Martin (Iowa State U.)

The edit distance of graphs

31 August 2010 9 / 23
A **hereditary property**, $\mathcal{H}$, of graphs is one that holds under 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.
A **hereditary property**, \( H \), of graphs is one that holds under the deletion of vertices.

I.e., if \( G \in H \), then every induced subgraph of \( G \) is in \( H \).

Examples:
- Planar graphs.
- Line graphs of bipartite graphs.
- Chordal graphs.
A **hereditary property**, $\mathcal{H}$, of graphs is one that holds under 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 $\geq 4$. 
A **HEREDITARY PROPERTY**, $\mathcal{H}$, of graphs is one that holds under 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 $\geq 4$.
- Perfect graphs
A hereditary property, \( \mathcal{H} \), of graphs is one that holds under 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 \( \geq 4 \).
- Perfect graphs: \( \chi = \omega \) for all induced subgraphs.
A **HEREDITARY PROPERTY**, $\mathcal{H}$, of graphs is one that holds under 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 $\geq 4$.
- Perfect graphs: $\chi = \omega$ for all induced subgraphs.
- $\text{Forb}(H)$
A hereditary property, $\mathcal{H}$, of graphs is one that holds under 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 $\geq 4$.
- Perfect graphs: $\chi = \omega$ for all induced subgraphs.
- $\text{Forb}(H)$: graphs with no induced copy of $H$. 
Definition

A **HEREDITARY PROPERTY**, \(H\), of graphs is one that holds under the deletion of vertices.

I.e., if \(G \in H\), then every induced subgraph of \(G\) is in \(H\).

Examples:

- Planar graphs.
- Line graphs of bipartite graphs.
- Chordal graphs: graphs with no chordless cycle of length \(\geq 4\).
- 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 the deletion of vertices.

Examples:

- Planar graphs.
- Line graphs of bipartite graphs.
- Chordal graphs: graphs with no chordless cycle of length \( \geq 4 \).
- 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.,

\[ \mathcal{H} = \text{Forb}(H), \text{ for some graph } H. \]
A useful parameter

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

But $V(H)$ can be partitioned into ANY combination of $a + c + 1$ independent sets and cliques.
A useful parameter

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

But $V(H)$ can be partitioned into ANY combination of $a + c + 1$ independent sets and cliques.

The quantity $a + c + 1$ is the BINARY CHROMATIC NUMBER, $\chi_B(H)$. (It is also called the COLO(U)RING NUMBER, $\tau(H)$.)
A useful parameter

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

But $V(H)$ can be partitioned into ANY combination of $a + c + 1$ independent sets and cliques.

The quantity $a + c + 1$ is the BINARY CHROMATIC NUMBER, $\chi_B(H)$. (It is also called the COLO(U)RING NUMBER, $\tau(H)$.)

**Theorem (Axenovich-Kézdy-M., 2008)**

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

$$\text{Dist}(n, \text{Forb}(H)) \geq \frac{1}{2(\chi_B(H) - 1)} \binom{n}{2} - o(n^2).$$
A useful parameter

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

But $V(H)$ can be partitioned into ANY combination of $a + c + 1$ independent sets and cliques.

The quantity $a + c + 1$ is the **BINAR Y CHROMATIC NUMBER**, $\chi_B(H)$. (It is also called the **COLO(u)RING NUMBER**, $\tau(H)$.)

**Theorem (Axenovich-Kézdy-M., 2008)**

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

$$\text{Dist}(n, \text{Forb}(H)) \geq \frac{1}{2(\chi_B(H) - 1)} \binom{n}{2} - o(n^2).$$

Moreover, this holds with equality if $H$ is self-complementary.
An example: 5-cycle

We will compute $\chi_B(C_5)$, the 5-cycle.
An example: 5-cycle

We will compute $\chi_B(C_5)$, the 5-cycle.

Let us consider all $(a, c)$ such that $a + c = 3$:

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

We will compute $\chi_B(C_5)$, the 5-cycle.

Let us consider all $(a, c)$ such that $a + c = 3$:

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

We will compute $\chi_B(C_5)$, the 5-cycle.

Let us consider all $(a, c)$ such that $a + c = 3$:

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

We will compute $\chi_B(C_5)$, the 5-cycle.

Let us consider all $(a, c)$ such that $a + c = 3$:

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

We will compute $\chi_B(C_5)$, the 5-cycle.

Let us consider all $(a, c)$ such that $a + c = 3$:

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

So, $\chi_B(C_5) \leq 3$. 
An example: 5-cycle

We will compute $\chi_B(C_5)$, the 5-cycle.

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

But it cannot be partitioned into, say, 2 independent sets and 0 cliques.
An example: 5-cycle

We will compute $\chi_B(C_5)$, the 5-cycle.

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

But it cannot be partitioned into, say, 2 independent sets and 0 cliques.

So, $\chi_B(C_5) \geq 3$. 
An example: 5-cycle

We will compute $\chi_B(C_5)$, the 5-cycle.

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

But it cannot be partitioned into, say, 2 independent sets and 0 cliques.

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

$$\text{Dist}(n, \text{Forb}(C_5)) = \frac{1}{2(\chi_B(H) - 1)} \binom{n}{2} - o(n^2).$$
Edit distance for $C_5$

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

$$\text{Dist}(n, \text{Forb}(C_5)) = \frac{1}{2} \binom{n}{2} - o(n^2).$$
Since $\chi_B(C_5) = 3$, and $C_5$ is self-complementary, the theorem gives

$$\text{Dist}(n, \text{Forb}(C_5)) = \frac{1}{4} \binom{n}{2} - o(n^2).$$

This $d^* (H)$ is the quantity we want to compute. It's the proportion of pairs that need to be changed, in the worst case.
Edit distance for $C_5$

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

$$\text{Dist}(n, \text{Forb}(C_5)) = \frac{1}{4} \binom{n}{2} - o(n^2).$$

Normalize and take the limit:

$$\text{Dist}(n, \text{Forb}(H))/\binom{n}{2} = \frac{1}{4} - o(1).$$
Edit distance for $C_5$

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

$$\text{Dist}(n, \text{Forb}(C_5)) = \frac{1}{4} \binom{n}{2} - o(n^2).$$

Normalize and take the limit:

$$\lim_{n \to \infty} \frac{\text{Dist}(n, \text{Forb}(H))}{\binom{n}{2}} = \frac{1}{4}.$$

We denote

$$d^*(\mathcal{H}) \overset{\text{def}}{=} \lim_{n \to \infty} \frac{\text{Dist}(n, \mathcal{H})}{\binom{n}{2}}.$$
Edit distance for $C_5$

\[ d^*(\mathcal{H}) \overset{\text{def}}{=} \lim_{n \to \infty} \frac{\text{Dist}(n, \mathcal{H})}{\binom{n}{2}}. \]

This $d^*(\mathcal{H})$ is the quantity we want to compute.

It’s the proportion of pairs that need to be changed, in the worst case.
Understanding $d^*$

Let $G_{n,p}$ denote the Erdős-Rényi random graph:

I.e., there are $n$ vertices and each edge is present, independently, with probability $p$. 

Theorem (Alon-Stav, 2008)

For every hereditary property, $H$, there exists a $p^* = \text{p}^*(H) \in [0,1]$ such that $d^*(H) = \lim_{n \to \infty} \frac{\text{Dist}(G_n, p^*, H)}{n^2}$.

Theorem (Balogh-M., 2008)

For every hereditary property, $H$, and every $p \in [0,1]$, if $g_H(p) = \lim_{n \to \infty} \frac{\text{Dist}(G_n, p, H)}{n^2}$ then it is also true that $g_H(p) = \lim_{n \to \infty} \max \{ \text{Dist}(G, H) : |V(G)| = n, |E(G)| = p(n^2) \} / n^2$.

Roughly, the hardest density-$p$-graph to edit is the random graph.
Let $G_{n,p}$ denote the Erdős-Rényi random graph:

I.e., there are $n$ vertices and each edge is present, independently, with probability $p$.

**Theorem (Alon-Stav, 2008)**

For every hereditary property, $\mathcal{H}$, there exists a $p^* = p^*(\mathcal{H}) \in [0, 1]$ such that

$$d^*(\mathcal{H}) = \lim_{n \to \infty} \frac{\text{Dist} (G_{n,p^*}, \mathcal{H})}{\binom{n}{2}}.$$  

(The expression $\text{Dist} (G_{n,p^*}, \mathcal{H})$ is tightly concentrated about the mean, so the right-hand side is well-defined.)
Understanding $d^*$

Theorem (Alon-Stav, 2008)

For every hereditary property, $\mathcal{H}$, there exists a $p^* = p^*(\mathcal{H}) \in [0, 1]$ such that

$$d^*(\mathcal{H}) = \lim_{n \to \infty} \text{Dist} \left( G_n, p^*, \mathcal{H} \right) / \binom{n}{2}.$$ 

Theorem (Balogh-M., 2008)

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

$$g_{\mathcal{H}}(p) = \lim_{n \to \infty} \frac{\text{Dist} \left( G_n, p, \mathcal{H} \right)}{\binom{n}{2}}$$

then it is also true that

$$g_{\mathcal{H}}(p) = \lim_{n \to \infty} \max \left\{ \text{Dist} \left( G, \mathcal{H} \right) : |V(G)| = n, |E(G)| = p \binom{n}{2} \right\} / \binom{n}{2}.$$
Theorem (Balogh-M., 2008)

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

$$g_{\mathcal{H}}(p) = \lim_{n \to \infty} \frac{\text{Dist}(G_n, p, \mathcal{H})}{n^2}$$

then it is also true that

$$g_{\mathcal{H}}(p) = \lim_{n \to \infty} \max \left\{ \text{Dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = p \binom{n}{2} \right\} \div \binom{n}{2}.$$

Roughly, the hardest density-$p$ graph to edit is the random graph.
The Edit Distance Function

Properties of $g_{\text{Forb}(H)}(p)$

Continuous and concave down.

$g_{\text{Forb}(H)}(0) = g_{\text{Forb}(H)}(1) = 0$.

For any rational $r \in [0, 1]$, there is an $H$, such that $p^*(\text{Forb}(H)) = r$.

There is an irrational $q \in [0, 1]$ and an $H$, such that $p^*(\text{Forb}(H)) = q$.
The Edit Distance Function

Properties of $g_{\text{Forb}(H)}(p)$

- Continuous and concave down.
The Edit Distance Function

Properties of $g_{\text{Forb}}(H)(p)$

- Continuous and concave down.
- $g_{\text{Forb}}(H)\left(\frac{1}{2}\right) = \frac{1}{2(\chi_B(H) - 1)}$. 

**Theorem (Balogh-M., 2008)**

Ryan Martin (Iowa State U.)
## Properties of $g_{\text{Forb}}(H)(p)$

- Continuous and concave down.
- \[
g_{\text{Forb}}(H)\left(\frac{1}{2}\right) = \frac{1}{2(\chi_B(H) - 1)}.
\]
- Achieves its maximum $(p^*, d^*)$ for some $p^* \in [0, 1]$. 

---

**Theorem (Balogh-M., 2008)**

Ryan Martin (Iowa State U.)
The Edit Distance Function

Properties of $g_{Forb}(H)(p)$

- Continuous and concave down.
- $g_{Forb}(H)\left(\frac{1}{2}\right) = \frac{1}{2(\chi_B(H) - 1)}$.
- Achieves its maximum $(p^*, d^*)$ for some $p^* \in [0, 1]$.
- If $H$ is neither complete nor empty, then $g_{Forb}(H)(0) = g_{Forb}(H)(1) = 0$. 

For any rational $r \in [0, 1]$, there is an $H$, such that $p^*(Forb(H)) = r$.

There is an irrational $q \in [0, 1]$ and an $H$, such that $p^*(Forb(H)) = q$. 

Theorem (Balogh-M., 2008)
The Edit Distance Function

Properties of $g_{\text{Forb}(H)}(p)$

- Continuous and concave down.
- $g_{\text{Forb}(H)}\left(\frac{1}{2}\right) = \frac{1}{2(\chi_B(H) - 1)}$.
- Achieves its maximum $(p^*, d^*)$ for some $p^* \in [0, 1]$.
- If $H$ is neither complete nor empty, then $g_{\text{Forb}(H)}(0) = g_{\text{Forb}(H)}(1) = 0$.
- For any rational $r \in [0, 1]$, there is an $H$, such that $p^*(\text{Forb}(H)) = r$.

Theorem (Balogh-M., 2008)

\[ p^*(\text{Forb}(K_a + E_b)) = \frac{a-1}{a+b-1} \quad d^*(\text{Forb}(K_a + E_b)) = \frac{1}{a+b-1}. \]
The Edit Distance Function

Properties of $g_{\text{Forb}(H)}(p)$

- Continuous and concave down.
- $g_{\text{Forb}(H)}(1/2) = \frac{1}{2(\chi_B(H) - 1)}$.
- Achieves its maximum $(p^*, d^*)$ for some $p^* \in [0, 1]$.
- If $H$ is neither complete nor empty, then $g_{\text{Forb}(H)}(0) = g_{\text{Forb}(H)}(1) = 0$.
- For any rational $r \in [0, 1]$, there is an $H$, such that $p^*(\text{Forb}(H)) = r$.
- There is an irrational $q \in [0, 1]$ and an $H$, such that $p^*(\text{Forb}(H)) = q$.

Theorem (Balogh-M., 2008)

$$p^*(\text{Forb}(K_{3,3})) = \sqrt{2} - 1 \quad d^*(\text{Forb}(K_{3,3})) = 3 - 2\sqrt{2}.$$
Example: $C_t$

**Theorem (Marchant-Thomason, 2010)**

$$g_{\text{Forb}}(C_4)(p) = p(1 - p) \quad d^*_{\text{Forb}}(C_4) = \frac{1}{4}.$$
Example: $C_t$

Theorem (Marchant, 2010++)

$$g_{\text{Forb}}(C_5)(p) = \min \left\{ \frac{p}{2}, \frac{1 - p}{2} \right\} \quad d^*_{\text{Forb}}(C_5) = \frac{1}{4}. $$

$$0.0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1.0$$

$$0.05 \quad 0.10 \quad 0.15 \quad 0.20 \quad 0.25$$
Example: $C_t$

**Theorem (M., 2010+)**

$$g_{\text{Forb}(C_6)}(p) = \min \left\{ p(1 - p), \frac{1 - p}{2} \right\} \quad d^*_\text{Forb}(C_6) = \frac{1}{4}.$$
Example: $C_t$

**Theorem (Marchant, 2010++)**

$$g_{\text{Forb}}(C_7)(p) = \min \left\{ \frac{p}{2}, \frac{p(1-p)}{1+p}, \frac{1-p}{3} \right\} \quad d^*_\text{Forb}(C_7) = 3 - 2\sqrt{2}.$$
Example: $C_t$

**Theorem (M., 2010+)**

$$g_{\text{Forb}(C_8)}(p) = \min \left\{ \frac{p(1-p)}{1+p}, \frac{1-p}{3} \right\}$$

$$d^*_{\text{Forb}(C_8)} = 3 - 2\sqrt{2}.$$

![Graph](image-url)
Example: $C_t$

**Theorem (M., 2010+)**

\[ g_{\text{Forb}}(C_9)(p) = \min \left\{ \frac{p}{2}, \frac{1 - p}{4} \right\} \quad d^*_{\text{Forb}}(C_9) = \frac{1}{6}. \]
Example: $C_t$

**Theorem (M., 2010+)**

\[ g_{\text{Forb}}(C_{10})(p) = \min \left\{ \frac{p(1-p)}{1+2p}, \frac{1-p}{4} \right\} \quad d_{\text{Forb}}^{*}(C_{10}) = \frac{2 - \sqrt{3}}{2}. \]
Example: $H_9$

**Theorem (M., 2010+)**

$$g_{\text{Forb}}(H_9)(p) = \min \left\{ \frac{p}{3}, \frac{p}{1 + 4p}, \frac{1 - p}{2} \right\} \quad d^{*}_{\text{Forb}}(K_{2,3}) = \frac{7 - \sqrt{17}}{16}.$$
Example: $K_{2,t}$

Theorem (M.-McKay, 2010++)

$$g_{Forb}(K_{2,2})(p) = p(1 - p) \quad d^*_{Forb}(K_{2,2}) = \frac{1}{4}.$$
Example: $K_{2,t}$

**Theorem (M.-McKay, 2010++)**

$$g_{\text{Forb}}(K_{2,3})(p) = \min \left\{ p(1 - p), \frac{1 - p}{2} \right\}$$

$$d_{\text{Forb}}^*(K_{2,3}) = \frac{1}{4}.$$
Theorem (M.-McKay, 2010++)

$$g_{\text{Forb}}(K_{2,4})(p) = \min \left\{ p(1 - p), \frac{1 + 7p}{15}, \frac{1 - p}{3} \right\} \quad d_{\text{Forb}}^*(K_{2,4}) = \frac{2}{9}.$$
A CRG from an SRG

The special editing recipe (called a COLORED REGULARITY GRAPH (CRG)) is derived from a particular STRONGLY REGULAR GRAPH (SRG).

The SRG is the generalized quadrangle $GQ(2,2)$. It has the following properties:

- has 15 vertices,
- is regular of degree 6,
- each pair of adjacent vertices have exactly 1 common neighbor, and
- each pair of nonadjacent vertices have exactly 3 common neighbors.

This is a so-called $(15, 6, 1, 3)$-SRG.
A CRG from an SRG

The special editing recipe (called a **colored regularity graph (CRG)**) is derived from a particular **strongly regular graph (SRG)**.

The SRG is the generalized quadrangle $GQ(2,2)$. It has the following properties:

- has 15 vertices,
- is regular of degree 6,
- each pair of adjacent vertices have exactly 1 common neighbor, and
- each pair of nonadjacent vertices have exactly 3 common neighbors.
A CRG from an SRG

The special editing recipe (called a **COLORED REGULARITY GRAPH (CRG)**) is derived from a particular **STRONGLY REGULAR GRAPH (SRG)**.

The SRG is the generalized quadrangle $GQ(2, 2)$. It has the following properties:

- has 15 vertices,
- is regular of degree 6,
- each pair of adjacent vertices have exactly 1 common neighbor, and
- each pair of nonadjacent vertices have exactly 3 common neighbors.

This is a so-called $(15, 6, 1, 3)$-SRG.
The SRG is the generalized quadrangle GQ(2, 2). It has the following properties:

- has 15 vertices,
- is regular of degree 6,
- each pair of adjacent vertices have exactly 1 common neighbor, and
- each pair of nonadjacent vertices have exactly 3 common neighbors.

This is a so-called (15, 6, 1, 3)-SRG.
Upper bounds

Computing upper bounds for $g_{\text{Forb}}(H)(p)$ is easy:

- Begin with a **COLORED REGULARITY GRAPH (CRG)**, or a “recipe” that tells how to partition the vertices of the random graph, $G(n, p)$ and how to add/delete edges to have no induced $H$. 
Computing upper bounds for $g_{\text{Forb}}(H)(p)$ is easy:

- Begin with a **colored regularity graph (CRG)**, or a “recipe” that tells how to partition the vertices of the random graph, $G(n, p)$ and how to add/delete edges to have no induced $H$.
- Optimally weight the vertices of the **CRG**, depending on $p$, to minimize the number of edge-operations.
Computing upper bounds for $g_{\text{Forb}(H)}(p)$ is easy:

- Begin with a **colored regularity graph (CRG)**, or a “recipe” that tells how to partition the vertices of the random graph, $G(n,p)$ and how to add/delete edges to have no induced $H$.
- Optimally weight the vertices of the CRG, depending on $p$, to minimize the number of edge-operations.
- This gives an upper bound for $g$. 
Lower bounds continue to be difficult.

If one knew all of the good CRGs or recipes, then $g_{\text{Forb}(H)}(p)$ could be computed. It is an infimum of functions of such CRGs.
Computing lower bounds

Lower bounds continue to be difficult.

If one knew all of the good **CRGs** or recipes, then $g_{\text{Forb}(H)}(p)$ could be computed. It is an infimum of functions of such CRGs.

We use a technique we called “localization” which exploits the fact that the optimal weighting of vertices is a quadratic program.

$$\min \left\{ \bar{x}^T M \bar{x} : \bar{x}^T \mathbf{1} = 1, \bar{x} \geq \mathbf{0} \right\}.$$
Marchant and Thomason show that constructions used to solve the Zarankiewicz problem are necessary for the edit distance function for $K_{3,3}$. This is not true for $K_{2,2}$, $K_{2,3}$ or $K_{2,4}$. In work in progress with Tracy McKay, we are working on computing $g_{\text{Forb}}(K_{2,t}(p))$ for $t \geq 5$. The “Zarankiewicz effect” seems to happen for $t \geq 9$. With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs. Other metrics. Which other metrics or functions of hereditary properties behave this way? The cut norm? The entropy function? Stability. What happens when many recipes give the best result? Understanding graphs. Can we use this information to solve classical graph problems? Yes. Pr¨ omel and Steger used the same approach to count graphs in a hereditary property. Graph limits, as studied by Borges, Chayes, Elek, Lov´ asz, B. Szegedy, Vesztergombi, et al.
Marchant and Thomason show that constructions used to solve the Zarankiewicz problem are necessary for the edit distance function for $K_{3,3}$. This is not true for $K_{2,2}$, $K_{2,3}$ or $K_{2,4}$.

In work in progress with Tracy McKay, we are working on computing $g_{\text{Forb}(K_{2,t})}(p)$ for $t \geq 5$. The “Zarankiewicz effect” seems to happen for $t \geq 9$. 

With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.

Other metrics. Which other metrics or functions of hereditary properties behave this way?

The cut norm?
The entropy function?

Stability. What happens when many recipes give the best result?

Understanding graphs. Can we use this information to solve classical graph problems?

Yes. Prömel and Steger used the same approach to count graphs in a hereditary property.

Graph limits, as studied by Borges, Chayes, Elek, Lovász, B. Szegedy, Vesztergombi, et al.
Marchant and Thomason show that constructions used to solve the Zarankiewicz problem are necessary for the edit distance function for $K_{3,3}$. This is not true for $K_{2,2}$, $K_{2,3}$ or $K_{2,4}$.

In work in progress with Tracy McKay, we are working on computing $g_{\text{Forb}(K_{2,t})}(p)$ for $t \geq 5$. The “Zarankiewicz effect” seems to happen for $t \geq 9$.

With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.
What’s next?

- Marchant and Thomason show that constructions used to solve the Zarankiewicz problem are necessary for the edit distance function for $K_{3,3}$. This is not true for $K_{2,2}$, $K_{2,3}$ or $K_{2,4}$.

- In work in progress with Tracy McKay, we are working on computing $g_{\text{Forb}(K_2,t)}(p)$ for $t \geq 5$. The “Zarankiewicz effect” seems to happen for $t \geq 9$.

- With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.

- Other metrics. Which other metrics or functions of hereditary properties behave this way?
What’s next?

- Marchant and Thomason show that constructions used to solve the Zarankiewicz problem are necessary for the edit distance function for $K_{3,3}$. This is not true for $K_{2,2}$, $K_{2,3}$ or $K_{2,4}$.

- In work in progress with Tracy McKay, we are working on computing $g_{\text{Forb}(K_{2,t})}(p)$ for $t \geq 5$. The “Zarankiewicz effect” seems to happen for $t \geq 9$.

- With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.

- Other metrics. Which other metrics or functions of hereditary properties behave this way?
  - The cut norm?
Marchant and Thomason show that constructions used to solve the Zarankiewicz problem are necessary for the edit distance function for $K_{3,3}$. This is not true for $K_{2,2}$, $K_{2,3}$ or $K_{2,4}$.

In work in progress with Tracy McKay, we are working on computing $g_{\text{Forb}(K_2,t)}(p)$ for $t \geq 5$. The “Zarankiewicz effect” seems to happen for $t \geq 9$.

With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.

Other metrics. Which other metrics or functions of hereditary properties behave this way?
  - The cut norm?
  - The entropy function?
In work in progress with Tracy McKay, we are working on computing $g_{\text{Forb}(K_2,t)}(p)$ for $t \geq 5$. The “Zarankiewicz effect” seems to happen for $t \geq 9$.

With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.

Other metrics. Which other metrics or functions of hereditary properties behave this way?
  - The cut norm?
  - The entropy function?

Stability. What happens when many recipes give the best result?
In work in progress with Tracy McKay, we are working on computing $g_{\text{Forb}(K_2, t)}(p)$ for $t \geq 5$. The “Zarankiewicz effect” seems to happen for $t \geq 9$.

With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.

Other metrics. Which other metrics or functions of hereditary properties behave this way?
- The cut norm?
- The entropy function?

Stability. What happens when many recipes give the best result?

Understanding graphs. Can we use this information to solve classical graph problems?
What’s next?

- With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.

- Other metrics. Which other metrics or functions of hereditary properties behave this way?
  - The cut norm?
  - The entropy function?

- Stability. What happens when many recipes give the best result?

- Understanding graphs. Can we use this information to solve classical graph problems?
  - Yes. Prömel and Steger used the same approach to count graphs in a hereditary property.
What’s next?

- With Maria Axenovich, we are looking at similar questions for multicolorings of complete graphs and directed graphs.

- Other metrics. Which other metrics or functions of hereditary properties behave this way?
  - The cut norm?
  - The entropy function?

- Stability. What happens when many recipes give the best result?

- Understanding graphs. Can we use this information to solve classical graph problems?
  - Yes. Prömel and Steger used the same approach to count graphs in a hereditary property.

- Graph limits, as studied by Borges, Chayes, Elek, Lovász, B. Szegedy, Vesztergombi, et al.
An unusual conjecture

What if the graph $H$ we want to edit away is a random graph, $G(n_0, p_0)$?

Formally, Fix $p_0 \in [0, 1]$ and let $H \sim G(n_0, p_0)$ with $\mathcal{H} = \text{Forb}(H)$. Then, with prob. $\to 1$ as $n_0 \to \infty$,

$$g_{\mathcal{H}}(p) = \min \left\{ \frac{2 \log_2 n_0}{n_0 \log_2 \frac{1}{1 - p_0}} p, \frac{2 \log_2 n_0}{n_0 \log_2 \frac{1}{p_0}} (1 - p) \right\} \pm o(1).$$

Alon and Stav verified this for $p_0 = 1/2$, yielding $p^* = 1/2$. I.e., it would be harder to edit $G(n_0, 0.25)$ out of a $G(n_0, 0.172)$ than a $G(n_0, 1/2)$.
An unusual conjecture

What if the graph $H$ we want to edit away is a random graph, $G(n_0, p_0)$?

Formally,

**Conjecture**

Fix $p_0 \in [0, 1]$ and let $H \sim G(n_0, p_0)$ with $\mathcal{H} = \text{Forb}(H)$. Then, with prob. $\to 1$ as $n_0 \to \infty$,

$$g_{\mathcal{H}}(p) = \min \left\{ \frac{2 \log_2 n_0}{n_0 \log_2 \frac{1}{1-p_0}} p, \frac{2 \log_2 n_0}{n_0 \log_2 \frac{1}{p_0}} (1 - p) \right\} \pm o(1).$$

Alon and Stav verified this for $p_0 = 1/2$, yielding $p^* = 1/2$. 
An unusual conjecture

What if the graph $H$ we want to edit away is a random graph, $G(n_0, p_0)$?

Formally,

**Conjecture**

Fix $p_0 \in [0, 1]$ and let $H \sim G(n_0, p_0)$ with $\mathcal{H} = \text{Forb}(H)$. Then, with prob. $\to 1$ as $n_0 \to \infty$,

$$g_{\mathcal{H}}(p) = \min \left\{ \frac{2 \log_2 n_0}{n_0 \log_2 \frac{1}{1-p_0}} p, \quad \frac{2 \log_2 n_0}{n_0 \log_2 \frac{1}{p_0}} (1-p) \right\} \pm o(1).$$

Alon and Stav verified this for $p_0 = 1/2$, yielding $p^* = 1/2$.

If the conjecture is false, it suggests unexpected behavior of $G(n, p)$. If the conjecture is true, it would imply $p^* \sim \frac{\log(1-p_0)}{\log(p_0(1-p_0))}$. 
An unusual conjecture

Formally,

**Conjecture**

Fix $p_0 \in [0, 1]$ and let $H \sim G(n_0, p_0)$ with $\mathcal{H} = \text{Forb}(H)$. Then, with prob. $\to 1$ as $n_0 \to \infty$,

$$g_{\mathcal{H}}(p) = \min \left\{ \frac{2 \log_2 n_0}{n_0 \log_2 \frac{1}{1-p_0}} p, \quad \frac{2 \log_2 n_0}{n_0 \log_2 \frac{1}{p_0}} (1-p) \right\} \pm o(1).$$

Alon and Stav verified this for $p_0 = 1/2$, yielding $p^* = 1/2$.

If the conjecture is false, it suggests unexpected behavior of $G(n, p)$.
If the conjecture is true, it would imply $p^* \sim \frac{\log(1-p_0)}{\log(p_0(1-p_0))}$.

I.e., it would be harder to edit $G(n_0, 0.25)$ out of a $G \sim G(n, 0.172)$ than a $G \sim G(n, 0.25)$.
My home page:

http://orion.math.iastate.edu/rymartin

My CV (with links to this and previous talks):

http://orion.math.iastate.edu/rymartin/cv/cv.pdf

Contact me:

rymartin.math@gmail.com