

# The edit distance on graphs, Part II

Ryan Martin

Iowa State University

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

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

- 1 Hereditary Properties and CRGs
- 2 Editing Procedure
- 3 Cores
- 4 Gray degrees
- 5 Example: Petersen  $\gamma$  function
- 6 Interesting CRG examples

# Structure

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

# Structure

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

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

For most of this talk, our hereditary properties are **PRINCIPAL**; i.e.,  $\mathcal{H} = \text{Forb}(H)$ , for some graph  $H$ .

# Structure

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

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

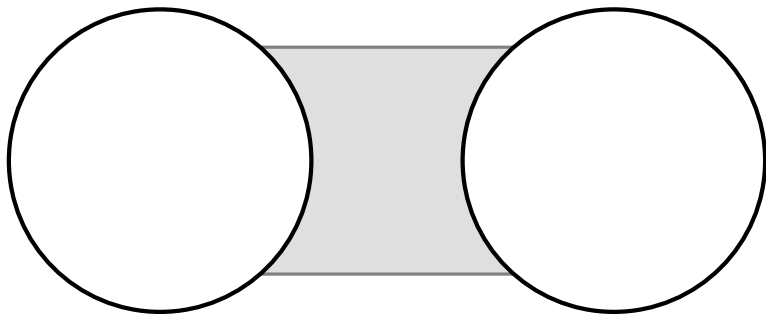
For most of this talk, our hereditary properties are **PRINCIPAL**; i.e.,  $\mathcal{H} = \text{Forb}(H)$ , for some graph  $H$ .

Let's revisit  $\text{Forb}(C_5)$ .

# Structure

Let's revisit  $\text{Forb}(C_5)$ .

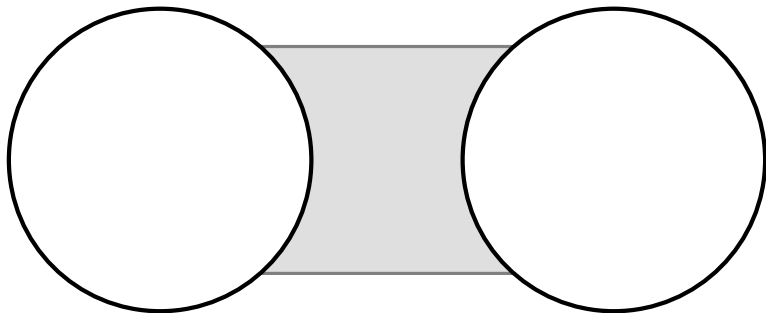
We can eliminate induced copies of  $C_5$  by partitioning the vertices in half and deleting the edges in each part.



# Structure

Let's revisit  $\text{Forb}(C_5)$ .

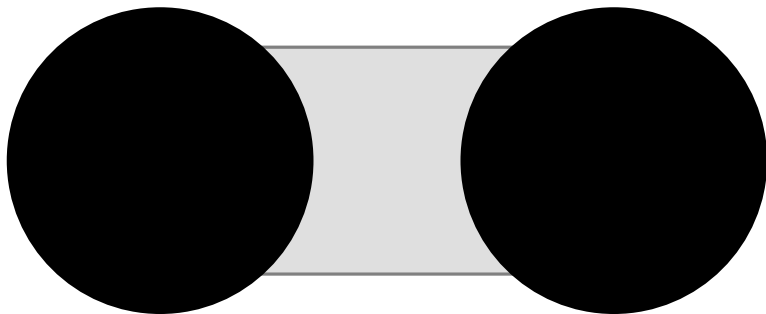
We can eliminate induced copies of  $C_5$  by partitioning the vertices in half *randomly* and deleting the edges in each part.



# Structure

Let's revisit  $\text{Forb}(C_5)$ .

We can eliminate induced copies of  $C_5$  by partitioning the vertices in half and adding the non-edges in each part.

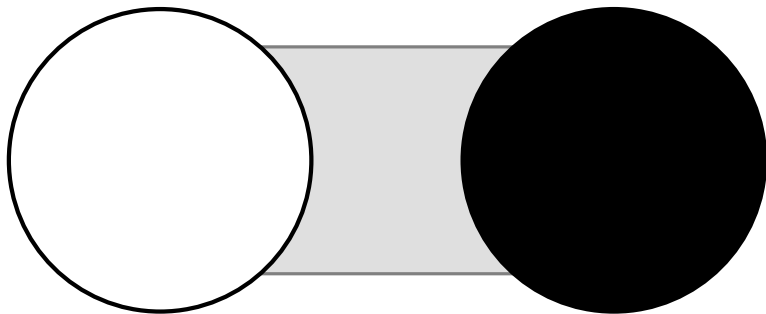




# Structure

Let's revisit  $\text{Forb}(C_5)$ .

We can eliminate induced copies of  $C_5$  by partitioning the vertices in half and deleting the edges in one part and adding the non-edges in the other part.



## Definition

A **COLORED REGULARITY GRAPH (CRG)**,  $K = (V, E)$ ,

is a complete graph such that

- $V(K) = VW \cup VB$  (vertices are white and black)
- $E(K) = EW \cup EG \cup EB$  (edges are white, gray, and black)

## Definition

A **COLORED REGULARITY GRAPH (CRG)**,  $K = (V, E)$ ,

is a complete graph such that

- $V(K) = VW \cup VB$  (vertices are white and black)
- $E(K) = EW \cup EG \cup EB$  (edges are white, gray, and black)

## Definition

A **COLORED HOMOMORPHISM**,  $\varphi : V(H) \rightarrow V(K)$ ,

is a map from a graph  $H = (V, E)$  to a CRG

$K = (VW, VB; EW, EG, EB)$  such that

- $vw \in E(\overline{H}) \implies \varphi(v) = \varphi(w) \in VW$  or  $\varphi(v)\varphi(w) \in EW \cup EG$ .
- $vw \in E(H) \implies \varphi(v) = \varphi(w) \in VB$  or  $\varphi(v)\varphi(w) \in EB \cup EG$ .

If there exists a colored homomorphism from  $H$  to  $K$ , then  $H \mapsto K$ .

# When there is no map

If

- $H$  is a graph,
- $K$  is a CRG, and
- $H \not\rightarrow K$ , then

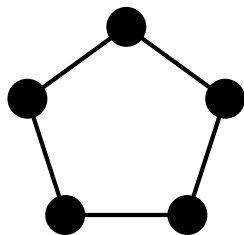
a graph  $G$  that “looks like  $K$ ” has no induced copy of  $H$ .

# When there is no map

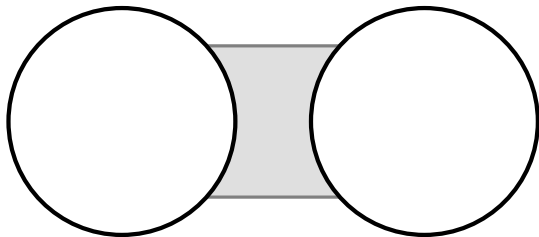
If

- $H$  is a graph,
- $K$  is a CRG, and
- $H \not\mapsto K$ , then

a graph  $G$  that “looks like  $K$ ” has no induced copy of  $H$ .



$G = C_5$



$K = K(2,0)$

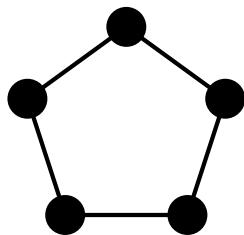
A graph that is bipartite has no induced  $C_5$ .

# When there is no map

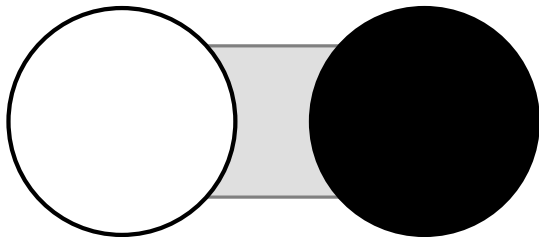
If

- $H$  is a graph,
- $K$  is a CRG, and
- $H \not\mapsto K$ , then

a graph  $G$  that “looks like  $K$ ” has no induced copy of  $H$ .



$G = C_5$



$K = K(1,1)$

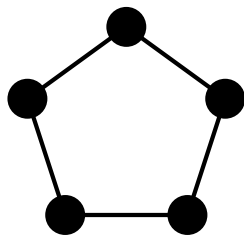
A graph that is split has no induced  $C_5$ .

# When there is no map

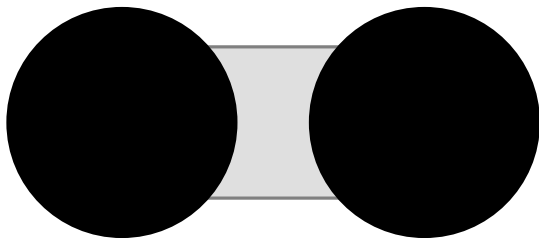
If

- $H$  is a graph,
- $K$  is a CRG, and
- $H \not\mapsto K$ , then

a graph  $G$  that “looks like  $K$ ” has no induced copy of  $H$ .



$G = C_5$



$K = K(0,2)$

A graph that is the complement of a bipartite graph has no induced  $C_5$ .

# When there is a map



# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .

- 1 Find a CRG  $K = (VW, VB; EW, EG, EB)$  such that  $H \not\rightarrow K$ .

# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .

- 1 Find a CRG  $K = (VW, VB; EW, EG, EB)$  such that  $H \not\rightarrow K$ .
- 2 Randomly map  $v \in V(G)$  to  $\varphi(v) \in V(K)$ .

# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .

- 1 Find a CRG  $K = (VW, VB; EW, EG, EB)$  such that  $H \not\rightarrow K$ .
- 2 Randomly map  $v \in V(G)$  to  $\varphi(v) \in V(K)$ .
- 3 If  $\varphi(v) = \varphi(w) \in VW$  or  $\varphi(v)\varphi(w) \in EW$  and  $vw \in E(G)$ , then delete  $vw$ .

# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .

- 1 Find a CRG  $K = (VW, VB; EW, EG, EB)$  such that  $H \not\rightarrow K$ .
- 2 Randomly map  $v \in V(G)$  to  $\varphi(v) \in V(K)$ .
- 3 If  $\varphi(v) = \varphi(w) \in VW$  or  $\varphi(v)\varphi(w) \in EW$  and  $vw \in E(G)$ , then  
delete  $vw$ .
- 4 If  $\varphi(v) = \varphi(w) \in VB$  or  $\varphi(v)\varphi(w) \in EB$  and  $vw \in E(\overline{G})$ , then  
add  $vw$ .

How many changes were made, on average? ( $k = |V(K)|$ )

# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .

- 1 Find a CRG  $K = (VW, VB; EW, EG, EB)$  such that  $H \not\rightarrow K$ .
- 2 Randomly map  $v \in V(G)$  to  $\varphi(v) \in V(K)$ .
- 3 If  $\varphi(v) = \varphi(w) \in VW$  or  $\varphi(v)\varphi(w) \in EW$  and  $vw \in E(G)$ , then  
delete  $vw$ .
- 4 If  $\varphi(v) = \varphi(w) \in VB$  or  $\varphi(v)\varphi(w) \in EB$  and  $vw \in E(\overline{G})$ , then  
add  $vw$ .

How many changes were made, on average? ( $k = |V(K)|$ )

$$\begin{aligned} \mathbb{E}[\# \text{ of changes}] &= \sum_{vw \in E(G)} \Pr\{\varphi(v)\varphi(w) \in (VW \cup EW)\} \\ &\quad + \sum_{vw \in E(\overline{G})} \Pr\{\varphi(v)\varphi(w) \in (VB \cup EB)\} \end{aligned}$$

# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .

- 1 Find a CRG  $K = (VW, VB; EW, EG, EB)$  such that  $H \not\rightarrow K$ .
- 2 Randomly map  $v \in V(G)$  to  $\varphi(v) \in V(K)$ .

How many changes were made, on average? ( $k = |V(K)|$ )

$$\begin{aligned}\mathbb{E}[\# \text{ of changes}] &= \sum_{vw \in E(G)} \Pr\{\varphi(v)\varphi(w) \in (VW \cup EW)\} \\ &\quad + \sum_{vw \in E(\bar{G})} \Pr\{\varphi(v)\varphi(w) \in (VB \cup EB)\} \\ &= \sum_{vw \in E(G)} \frac{|VW| + 2|EW|}{k^2} + \sum_{vw \in E(\bar{G})} \frac{|VB| + 2|EB|}{k^2}\end{aligned}$$

# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .

- 1 Find a CRG  $K = (VW, VB; EW, EG, EB)$  such that  $H \not\vdash K$ .
- 2 Randomly map  $v \in V(G)$  to  $\varphi(v) \in V(K)$ .

How many changes were made, on average? ( $k = |V(K)|$ )

$$\begin{aligned}\mathbb{E}[\# \text{ of changes}] &= \sum_{vw \in E(G)} \Pr\{\varphi(v)\varphi(w) \in (VW \cup EW)\} \\ &\quad + \sum_{vw \in E(\bar{G})} \Pr\{\varphi(v)\varphi(w) \in (VB \cup EB)\} \\ &= \sum_{vw \in E(G)} \frac{|VW| + 2|EW|}{k^2} + \sum_{vw \in E(\bar{G})} \frac{|VB| + 2|EB|}{k^2} \\ &= \frac{|VW| + 2|EW|}{k^2} p \binom{n}{2} + \frac{|VB| + 2|EB|}{k^2} (1-p) \binom{n}{2}\end{aligned}$$

# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .

- 1 Find a CRG  $K = (VW, VB; EW, EG, EB)$  such that  $H \not\vdash K$ .
- 2 Randomly map  $v \in V(G)$  to  $\varphi(v) \in V(K)$ .

How many changes were made, on average? ( $k = |V(K)|$ )

$$\begin{aligned}\mathbb{E}[\# \text{ of changes}] &= \sum_{vw \in E(G)} \Pr\{\varphi(v)\varphi(w) \in (VW \cup EW)\} \\ &\quad + \sum_{vw \in E(\overline{G})} \Pr\{\varphi(v)\varphi(w) \in (VB \cup EB)\} \\ &= \frac{|VW| + 2|EW|}{k^2} p \binom{n}{2} + \frac{|VB| + 2|EB|}{k^2} (1-p) \binom{n}{2}\end{aligned}$$

$$\text{dist}(G, \text{Forb}(H)) \leq \frac{1}{k^2} [p(|VW| + 2|EW|) + (1-p)(|VB| + 2|EB|)]$$



# How to edit

Let  $G$  be a density- $\rho$  graph. Let us transform it into a  $G' \in \text{Forb}(H)$ .  
How many changes were made, on average? ( $k = |V(K)|$ )

$$\text{dist}(G, \text{Forb}(H)) \leq \inf_K \left\{ \frac{1}{k^2} [\rho(|VW| + 2|EW|) + (1-\rho)(|VB| + 2|EB|)] \right\}$$

The infimum is taken over all  $K$  such that  $H \not\rightarrow K$ .

We denote

$$\mathcal{K}(\text{Forb}(H)) := \{K : H \not\rightarrow K\}.$$

# How to edit

Let  $G$  be a density- $p$  graph. Let us transform it into a  $G' \in \mathcal{H}$ .  
How many changes were made, on average? ( $k = |V(K)|$ )

$$\text{dist}(G, \mathcal{H}) \leq \inf_K \left\{ \frac{1}{k^2} [p(|VW| + 2|EW|) + (1-p)(|VB| + 2|EB|)] \right\}$$

The infimum is taken over all  $K$  such that  $H \not\rightarrow K, \forall H \in \mathcal{F}(\mathcal{H})$ .  
We denote

$$\mathcal{K}(\mathcal{H}) := \{K : H \not\rightarrow K, \forall H \in \mathcal{F}(\mathcal{H})\}.$$

where  $\mathcal{H} = \bigcap \{\text{Forb}(H) : H \in \mathcal{F}(\mathcal{H})\}$ .

# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

Consider the Erdős-Rényi random graph model  $G(n, p)$ .

The graph has  $n$  vertices and each pair has an edge with probability  $p$ .

# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

Consider the Erdős-Rényi random graph model  $G(n, p)$ .

The graph has  $n$  vertices and each pair has an edge with probability  $p$ .

- Choose  $G \sim G(n, p)$ .

# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

Consider the Erdős-Rényi random graph model  $G(n, p)$ .

The graph has  $n$  vertices and each pair has an edge with probability  $p$ .

- Choose  $G \sim G(n, p)$ .
- Let  $G' \in \mathcal{H}$  be the closest graph to  $G$ .

# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

Consider the Erdős-Rényi random graph model  $G(n, p)$ .

The graph has  $n$  vertices and each pair has an edge with probability  $p$ .

- Choose  $G \sim G(n, p)$ .
- Let  $G' \in \mathcal{H}$  be the closest graph to  $G$ .
- Apply Szemerédi's Regularity Lemma *to*  $G'$  with  $\epsilon$ .

# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

Consider the Erdős-Rényi random graph model  $G(n, p)$ .

The graph has  $n$  vertices and each pair has an edge with probability  $p$ .

- Choose  $G \sim G(n, p)$ .
- Let  $G' \in \mathcal{H}$  be the closest graph to  $G$ .
- Apply Szemerédi's Regularity Lemma *to  $G'$*  with  $\epsilon$ .
- We need to apply it twice – once to  $G'$  and once inside each cluster.  
A version of RegLem due to Alon, Fischer, Krivelevich, and M. Szegedy [2000] does this efficiently.



# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

Consider the Erdős-Rényi random graph model  $G(n, p)$ .

The graph has  $n$  vertices and each pair has an edge with probability  $p$ .

- Choose  $G \sim G(n, p)$ .
- Let  $G' \in \mathcal{H}$  be the closest graph to  $G$ .
- Apply Szemerédi's Regularity Lemma *to  $G'$*  with  $\epsilon$ .
- Edit the very dense and very sparse regular pairs and clusters to look like  $K \in \mathcal{K}(\mathcal{H})$ .

# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

Consider the Erdős-Rényi random graph model  $G(n, p)$ .

The graph has  $n$  vertices and each pair has an edge with probability  $p$ .

- Choose  $G \sim G(n, p)$ .
- Let  $G' \in \mathcal{H}$  be the closest graph to  $G$ .
- Apply Szemerédi's Regularity Lemma *to  $G'$*  with  $\epsilon$ .
- Edit the very dense and very sparse regular pairs and clusters to look like  $K \in \mathcal{K}(\mathcal{H})$ .
- # of changes is  $\geq f_K(p) \binom{n}{2} - \frac{\epsilon}{2} n^2$  with high probability.

# Best possible

For any CRG,  $K = (VW, VB; EW, EG, EB)$ , denote

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

If  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , then

$$\text{dist}(G, \mathcal{H}) \leq \inf \{f_K(p) : K \in \mathcal{K}(\mathcal{H})\}.$$

- Choose  $G \sim G(n, p)$ .
- Let  $G' \in \mathcal{H}$  be the closest graph to  $G$ .
- # of changes is  $\geq f_K(p) \binom{n}{2} - \frac{\epsilon}{2} n^2$  with high probability.

For every  $\epsilon > 0$ , there exists a  $\mathcal{K}(\mathcal{H}) = \{K : H \not\rightarrow K\}$ , such that with high probability,

$$\text{dist}(G(n, p), \mathcal{H}) \geq f_K(p) - \epsilon.$$

# The theorem revisited

## Theorem (Balogh-M., 2008)

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

$$\begin{aligned} \text{ed}_{\mathcal{H}}(p) &= \lim_{n \rightarrow \infty} \mathbb{E}[\text{dist}(G(n, p), \mathcal{H})] \\ &= \lim_{n \rightarrow \infty} \max \{ \text{dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p \binom{n}{2} \rfloor \} \end{aligned}$$

# The theorem revisited

## Theorem (Balogh-M., 2008)

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

$$\begin{aligned} \text{ed}_{\mathcal{H}}(p) &= \lim_{n \rightarrow \infty} \mathbb{E}[\text{dist}(G(n, p), \mathcal{H})] \\ &= \lim_{n \rightarrow \infty} \max \{ \text{dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p \binom{n}{2} \rfloor \} \\ &= \inf \{ f_K(p) : K \in \mathcal{K}(\mathcal{H}) \}, \end{aligned}$$

where

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

# The theorem revisited

## Theorem (Balogh-M., 2008)

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

$$\begin{aligned} \text{ed}_{\mathcal{H}}(p) &= \lim_{n \rightarrow \infty} \mathbb{E}[\text{dist}(G(n, p), \mathcal{H})] \\ &= \lim_{n \rightarrow \infty} \max \{ \text{dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p \binom{n}{2} \rfloor \} \\ &= \inf \{ f_K(p) : K \in \mathcal{K}(\mathcal{H}) \}, \end{aligned}$$

where

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

So we only need to check all CRGs in  $\mathcal{K}(\mathcal{H})$ .

# The theorem revisited

## Theorem (Balogh-M., 2008)

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

$$\begin{aligned} \text{ed}_{\mathcal{H}}(p) &= \lim_{n \rightarrow \infty} \mathbb{E}[\text{dist}(G(n, p), \mathcal{H})] \\ &= \lim_{n \rightarrow \infty} \max \{ \text{dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p \binom{n}{2} \rfloor \} \\ &= \inf \{ f_K(p) : K \in \mathcal{K}(\mathcal{H}) \}, \end{aligned}$$

where

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)].$$

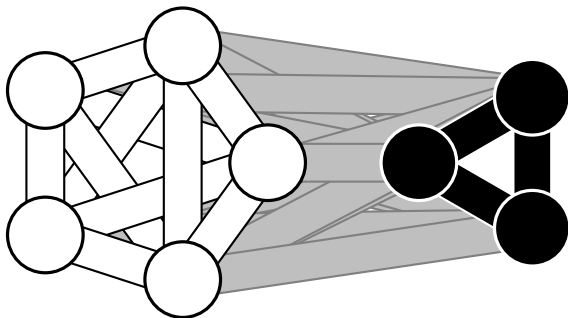
So we only need to check all CRGs in  $\mathcal{K}(\mathcal{H})$ .  $|\mathcal{K}(\mathcal{H})| = \infty$ .

But there are some important features of  $K \in \mathcal{K}(\mathcal{H})$  we can exploit.



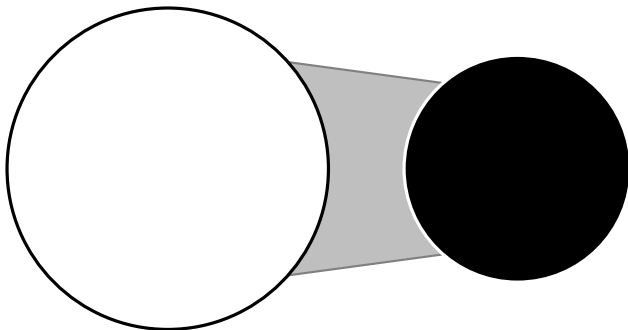
For this CRG,  $K$ , we have  $C_4 \not\rightarrow K$ .





For this CRG,  $K$ , we have  $C_4 \not\rightarrow K$ .

If  $p$  is small enough, then editing  $G(n, p)$  this way is more efficient than equal weights.



For this CRG,  $K$ , we have  $C_4 \not\rightarrow K$ .

If  $p$  is small enough, then editing  $G(n, p)$  this way is more efficient than equal weights.

Why “clone” vertices, when we can weight two vertices in an optimal way?

# Weighting

For this CRG,  $K$ , we have  $C_4 \not\rightarrow K$ .

If  $p$  is small enough, then editing  $G(n, p)$  this way is more efficient than equal weights.

Why “clone” vertices, when we can weight two vertices in an optimal way?

$$f_K(p) = \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)]$$

# Weighting

For this CRG,  $K$ , we have  $C_4 \not\rightarrow K$ .

If  $p$  is small enough, then editing  $G(n, p)$  this way is more efficient than equal weights.

Why “clone” vertices, when we can weight two vertices in an optimal way?

$$\begin{aligned} f_K(p) &= \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)] \\ &= \left(\frac{1}{k}\mathbf{1}\right)^T \mathbf{M} \left(\frac{1}{k}\mathbf{1}\right) \end{aligned}$$

where  $\mathbf{1}^T = [1, 1, \dots, 1]$  and

$$(\mathbf{M})_{ij} = \begin{cases} p, & \text{if } \{i, j\} \in VW \cup EW; \\ 1 - p, & \text{if } \{i, j\} \in VB \cup EB; \\ 0, & \text{else.} \end{cases}$$

# Weighting

For this CRG,  $K$ , we have  $C_4 \not\rightarrow K$ .

If  $p$  is small enough, then editing  $G(n, p)$  this way is more efficient than equal weights.

Why “clone” vertices, when we can weight two vertices in an optimal way?

$$f_K(p) = \left(\frac{1}{k}\mathbf{1}\right)^T \mathbf{M} \left(\frac{1}{k}\mathbf{1}\right)$$

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}$$

where  $\mathbf{1}^T = [1, 1, \dots, 1]$  and  $\mathbf{0}^T = [0, 0, \dots, 0]$  and

$$(\mathbf{M})_{ij} = \begin{cases} p, & \text{if } \{i, j\} \in VW \cup EW; \\ 1 - p, & \text{if } \{i, j\} \in VB \cup EB; \\ 0, & \text{else.} \end{cases}$$

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

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

$$\begin{aligned} \text{ed}_{\mathcal{H}}(p) &= \lim_{n \rightarrow \infty} \mathbb{E} [\text{dist}(G(n, p), \mathcal{H})] \\ &= \lim_{n \rightarrow \infty} \max \{ \text{dist}(G, \mathcal{H}) : |V(G)| = n, |E(G)| = \lfloor p \binom{n}{2} \rfloor \} \\ &= \inf \{ f_K(p) : K \in \mathcal{K}(\mathcal{H}) \} \\ &= \inf \{ g_K(p) : K \in \mathcal{K}(\mathcal{H}) \} \end{aligned}$$

where

$$\begin{aligned} f_K(p) &= \frac{1}{k^2} [p(|VW| + 2|EW|) + (1 - p)(|VB| + 2|EB|)] \\ g_K(p) &= \min \{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \} \end{aligned}$$

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

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

$$\text{ed}_{\mathcal{H}}(p) = \inf \{g_K(p) : K \in \mathcal{K}(\mathcal{H})\}$$

where

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}$$

Theorem (Balogh-M., 2008; Marchant-Thomason, 2010)

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

$$\text{ed}_{\mathcal{H}}(p) = \min \{g_K(p) : K \in \mathcal{K}(\mathcal{H})\}$$

where

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}$$

What if the optimal  $\mathbf{x}$  has a zero entry?



Theorem (Balogh-M., 2008; Marchant-Thomason, 2010)

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

$$\text{ed}_{\mathcal{H}}(p) = \min \{g_K(p) : K \in \mathcal{K}(\mathcal{H})\}$$

where

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}$$

What if the optimal  $\mathbf{x}$  has a zero entry? Might as well delete the vertex.

## Theorem (Balogh-M., 2008; Marchant-Thomason, 2010)

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

$$\text{ed}_{\mathcal{H}}(p) = \min \{g_K(p) : K \in \mathcal{K}(\mathcal{H})\}$$

where

$$g_K(p) = \min \{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \}$$

What if the optimal  $\mathbf{x}$  has a zero entry? Might as well delete the vertex.

## Definition

For  $p \in (0, 1)$ , a CRG  $K$  is called  **$p$ -CORE** if

$$g_K(p) < g_{K'}(p)$$

for any proper sub-CRG  $K'$ .

## Structure of $p$ -cores

Consider a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$

# Structure of $p$ -cores

Consider a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$

## Proposition

If the matrix  $\mathbf{M}$  has three entries and the optimal  $\mathbf{x}$  of the above quadratic program has no zero entries, then for  $i \neq j$ ,

$$m_{ij} < \min\{m_{ii}, m_{jj}\}.$$

# Structure of $p$ -cores

Consider a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$

## Proposition

If the matrix  $\mathbf{M}$  has three entries and the optimal  $\mathbf{x}$  of the above quadratic program has no zero entries, then for  $i \neq j$ ,

$$m_{ij} < \min\{m_{ii}, m_{jj}\}.$$

This allows us to categorize  $p$ -core CRGs:

## Theorem (Marchant-Thomason, 2010)

Let  $p \in [0, 1]$  and let  $K = (\text{VW}, \text{VB}; \text{EW}, \text{EG}, \text{EB})$  be a  $p$ -core CRG.

- Ⓐ  $p \leq 1/2 \implies \text{EB} = \emptyset$  and no EW is incident to VW.
- Ⓑ  $p \geq 1/2 \implies \text{EW} = \emptyset$  and no EB is incident to VB.

# Structure of $p$ -cores

Consider a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$

This allows us to categorize  $p$ -core CRGs:

## Theorem (Marchant-Thomason, 2010)

Let  $p \in [0, 1]$  and let  $K = (\text{VW}, \text{VB}; \text{EW}, \text{EG}, \text{EB})$  be a  $p$ -core CRG.

- a)  $p \leq 1/2 \implies \text{EB} = \emptyset$  and no EW is incident to VW.
- b)  $p \geq 1/2 \implies \text{EW} = \emptyset$  and no EB is incident to VB.

In particular,  $1/2$ -core CRGs have only gray edges. Hence the result

$$\text{ed}_{\mathcal{H}} \left( \frac{1}{2} \right) = \frac{1}{2(\chi_{\text{B}}(\mathcal{H}) - 1)}.$$

# Gray degree

Consider a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$

## Gray degree

Consider ... again ... a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$



## Gray degree

Consider ... again ... a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$

### Proposition

If the matrix  $\mathbf{M}$  has nonnegative entries and the optimal  $\mathbf{x}$  of the above quadratic program has no zero entries, then

$$\mathbf{M} \mathbf{x} = g_K(p) \mathbf{1}.$$

## Gray degree

Consider ... again ... a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$

### Proposition

If the matrix  $\mathbf{M}$  has nonnegative entries and the optimal  $\mathbf{x}$  of the above quadratic program has no zero entries, then

$$\mathbf{M} \mathbf{x} = g_K(p) \mathbf{1}.$$

This allows us to determine the value of the “gray degree” of each vertex:

# Gray degree

Consider ... again ... a  $p$ -core CRG  $K$  and the following quadratic program:

$$g_K(p) = \min \left\{ \mathbf{x}^T \mathbf{M} \mathbf{x} : \mathbf{x}^T \mathbf{1} = 1, \mathbf{x} \geq \mathbf{0} \right\}.$$

This allows us to determine the value of the “gray degree” of each vertex:

## Theorem

*For each vertex,  $v$  in a CRG  $K$ , with  $g := g_K(p)$  as defined by the above quadratic program, we denote  $\mathbf{x}(v)$  to be the weight of  $v$  in the optimal  $\mathbf{x}$  and  $d_G(v)$  the sum of the weights of the gray neighbors of  $v$ .*

- 1  $p \leq 1/2 \implies d_G(v) = \frac{p-g}{p} + \frac{1-2p}{p} \mathbf{x}(v), \quad \forall v \in \text{VB}.$
- 2  $p \geq 1/2 \implies d_G(v) = \frac{1-p-g}{1-p} + \frac{2p-1}{1-p} \mathbf{x}(v), \quad \forall v \in \text{VW}.$

# Gray degree

This allows us to determine the value of the “gray degree” of each vertex:

## Theorem

$$\textcircled{1} \quad p \leq 1/2 \implies d_G(v) = \frac{p-g}{p} + \frac{1-2p}{p} \mathbf{x}(v), \quad \forall v \in VB.$$

$$\textcircled{2} \quad p \geq 1/2 \implies d_G(v) = \frac{1-p-g}{1-p} + \frac{2p-1}{1-p} \mathbf{x}(v), \quad \forall v \in VW.$$

## Corollary

$$\textcircled{1} \quad p \leq 1/2 \implies \mathbf{x}(v) \leq \frac{g}{1-p}, \quad \forall v \in VB.$$

$$\textcircled{2} \quad p \geq 1/2 \implies \mathbf{x}(v) \leq \frac{g}{p}, \quad \forall v \in VW.$$

# Archipelago theorem

## Definition

Let  $K$  be a CRG. A sub-CRG  $K'$  is a **COMPONENT** if,  $\forall v_1, v_2 \in V(K')$ , there exists a path in  $K'$  from  $v_1$  to  $v_2$  without using a gray edge.

# Archipelago theorem

## Definition

Let  $K$  be a CRG. A sub-CRG  $K'$  is a **COMPONENT** if,  $\forall v_1, v_2 \in V(K')$ , there exists a path in  $K'$  from  $v_1$  to  $v_2$  without using a gray edge. Consequently,  $vw \in EG, \forall v \in V(K')$  and  $\forall w \in V(K) - V(K')$ .

# Archipelago theorem

## Definition

Let  $K$  be a CRG. A sub-CRG  $K'$  is a **COMPONENT** if,  $\forall v_1, v_2 \in V(K')$ , there exists a path in  $K'$  from  $v_1$  to  $v_2$  without using a gray edge. Consequently,  $vw \in EG, \forall v \in V(K')$  and  $\forall w \in V(K) - V(K')$ .

## Theorem

If  $K$  is a CRG with components  $K_1, \dots, K_\ell$ , then

$$g_K^{-1}(p) = \sum_{i=1}^{\ell} g_{K_i}^{-1}(p).$$

# Archipelago theorem

## Definition

Let  $K$  be a CRG. A sub-CRG  $K'$  is a **COMPONENT** if,  $\forall v_1, v_2 \in V(K')$ , there exists a path in  $K'$  from  $v_1$  to  $v_2$  without using a gray edge. Consequently,  $vw \in EG, \forall v \in V(K')$  and  $\forall w \in V(K) - V(K')$ .

## Theorem

If  $K$  is a CRG with components  $K_1, \dots, K_\ell$ , then

$$g_K^{-1}(p) = \sum_{i=1}^{\ell} g_{K_i}^{-1}(p).$$

## Corollary

If  $K$  has only gray edges,  $r = |VW|$ , and  $s = |VB|$ , then it is  $K(r, s)$  and

$$g_{K(r,s)}(p) = \left( \frac{r}{p} + \frac{s}{1-p} \right)^{-1} = \frac{p(1-p)}{r(1-p) + sp}.$$



## Corollary

If  $K$  has only gray edges,  $r = |VW|$ , and  $s = |VB|$ , then it is  $K(r, s)$  and

$$g_{K(r,s)}(p) = \left( \frac{r}{p} + \frac{s}{1-p} \right)^{-1} = \frac{p(1-p)}{r(1-p) + sp}.$$

An upper bound for  $\text{ed}_{\mathcal{H}}(p)$  comes from considering only gray-edge CRGs.

$$\text{ed}_{\mathcal{H}}(p) = \min \{g_K(p) : K \not\rightarrow \mathcal{K}(\mathcal{H})\}$$

$$\gamma_{\mathcal{H}}(p) := \min \{g_K(p) : K \not\rightarrow \mathcal{K}(\mathcal{H}), E(K) = \text{EG}(K)\}$$

# Archipelago theorem

## Corollary

If  $K$  has only gray edges,  $r = |VW|$ , and  $s = |VB|$ , then it is  $K(r, s)$  and

$$g_{K(r,s)}(p) = \left( \frac{r}{p} + \frac{s}{1-p} \right)^{-1} = \frac{p(1-p)}{r(1-p) + sp}.$$

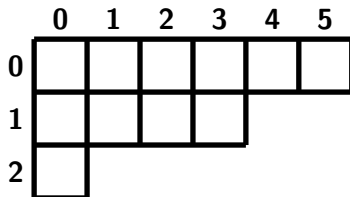
An upper bound for  $\text{ed}_{\mathcal{H}}(p)$  comes from considering only gray-edge CRGs.

$$\text{ed}_{\mathcal{H}}(p) = \min \{g_K(p) : K \not\rightarrow \mathcal{K}(\mathcal{H})\}$$

$$\gamma_{\mathcal{H}}(p) := \min \{g_K(p) : K \not\rightarrow \mathcal{K}(\mathcal{H}), E(K) = \text{EG}(K)\}$$

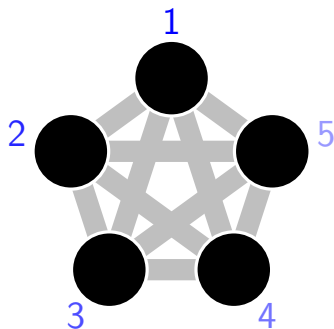
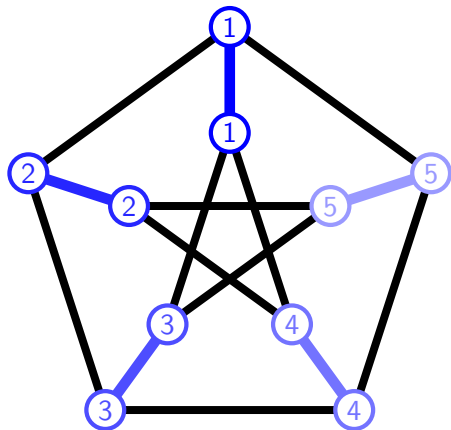
The set of  $(r, s)$  such that  $K(r, s) \in \mathcal{K}(\mathcal{H})$  forms the **CLIQUE SPECTRUM**, a Young tableau.

$\Gamma(\text{Forb}(C_{11})):$



$(0, 5) \notin \Gamma$

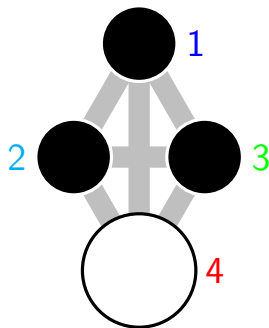
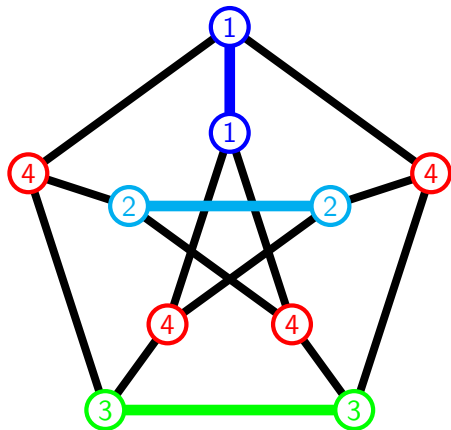
Let  $P_{10}$  be the Petersen graph.  
Let  $\Gamma(\text{Forb}(P_{10}))$  be the clique spectrum of  $\text{Forb}(P_{10})$ .



$(0, 5) \notin \Gamma(\text{Forb}(P_{10}))$

$$(1, 3) \notin \Gamma$$

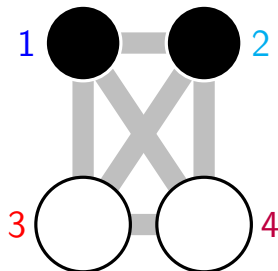
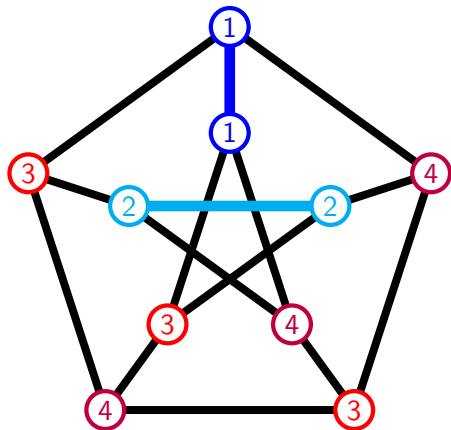
Let  $P_{10}$  be the Petersen graph.  
Let  $\Gamma(\text{Forb}(P_{10}))$  be the clique spectrum of  $\text{Forb}(P_{10})$ .



$$(1, 3) \notin \Gamma(\text{Forb}(P_{10}))$$

$(2, 2) \notin \Gamma$

Let  $P_{10}$  be the Petersen graph.  
Let  $\Gamma(\text{Forb}(P_{10}))$  be the clique spectrum of  $\text{Forb}(P_{10})$ .



$(2, 2) \notin \Gamma(\text{Forb}(P_{10}))$

# Clique spectrum

$\Gamma(\text{Forb}(P_{10})):$

	0	1	2	3	4
0					
1					
2					

$\Gamma(\text{Forb}(P_{10}))$ :

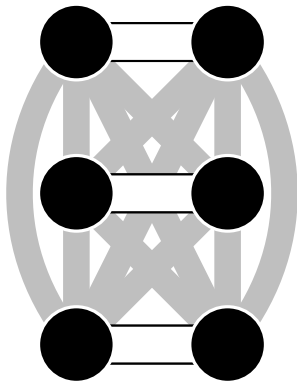
	0	1	2	3	4
0					
1					
2					

$$g_{K(2,1)}(p) = \left( \frac{2}{p} + \frac{1}{1-p} \right)^{-1} = \frac{p(1-p)}{2-p}$$

$$g_{K(1,2)}(p) = \left( \frac{1}{p} + \frac{2}{1-p} \right)^{-1} = \frac{p(1-p)}{1+p}$$

$$g_{K(0,4)}(p) = \left( \frac{0}{p} + \frac{4}{1-p} \right)^{-1} = \frac{1-p}{4}$$

$$\gamma_{\text{Forb}(P_{10})} = \min \left\{ \frac{p(1-p)}{2-p}, \frac{p(1-p)}{1+p}, \frac{1-p}{4} \right\}$$

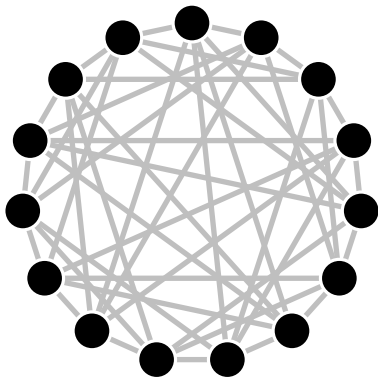


$$K_{2,5} \not\rightarrow K$$

$$g_K(p) = \frac{1}{6}, \text{ if } p < 1/2.$$



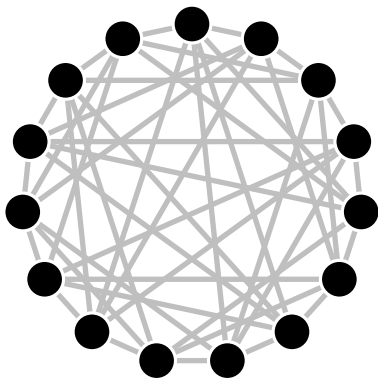
# Strongly-regular graphs



$$K_{2,4} \not\rightarrow K$$

$$g_K(p) = \frac{1+7p}{15}, \text{ if } p < 1/2.$$

# Strongly-regular graphs



t	=	4
k	=	15
d	=	6

$$K_{2,t} \not\rightarrow \text{SRG}(k, d, \leq t-3, \leq t-1)$$

$$g_K(p) = \frac{1+(k-d-2)p}{k}, \text{ if } p < 1/2.$$

# Zarankiewicz constructions

W.G. Brown constructed dense  $K_{3,3}$ -free graphs.

## Lemma

*If  $K$  is a CRG with  $V(K) = VB$  and  $E(K) = EW \cup EG$ , then  $K_{3,3} \not\rightarrow K$  iff  $K$  does not have a gray  $K_3$  or a gray  $K_{3,3}$ .*

# Zarankiewicz constructions

W.G. Brown constructed dense  $K_{3,3}$ -free graphs.

## Lemma

*If  $K$  is a CRG with  $V(K) = VB$  and  $E(K) = EW \cup EG$ , then  $K_{3,3} \not\rightarrow K$  iff  $K$  does not have a gray  $K_3$  or a gray  $K_{3,3}$ .*

- Given a (looped) graph  $G$  with no  $K_{3,3}$ , create a bipartite  $G'$  by cloning each vertex.
- For distinct  $x, y \in V(G)$ , let  $x_1 \sim_{G'} y_2$  iff  $x \sim_G y$ .
- For  $x \in V(G)$ , let  $x_1 \sim_{G'} x_2$  iff  $x$  has a loop.

# Zarankiewicz constructions

W.G. Brown constructed dense  $K_{3,3}$ -free graphs.

## Lemma

*If  $K$  is a CRG with  $V(K) = VB$  and  $E(K) = EW \cup EG$ , then  $K_{3,3} \not\rightarrow K$  iff  $K$  does not have a gray  $K_3$  or a gray  $K_{3,3}$ .*

- Given a (looped) graph  $G$  with no  $K_{3,3}$ , create a bipartite  $G'$  by cloning each vertex.
- For distinct  $x, y \in V(G)$ , let  $x_1 \sim_{G'} y_2$  iff  $x \sim_G y$ .
- For  $x \in V(G)$ , let  $x_1 \sim_{G'} x_2$  iff  $x$  has a loop.
- If  $G$  has  $m$  edges (loops count  $1/2$ ), then

$$f_K(p) = \frac{1 - 2p}{2n} + p - \frac{pm}{n^2}.$$

# Zarankiewicz constructions

W.G. Brown constructed dense  $K_{3,3}$ -free graphs.

## Lemma

If  $K$  is a CRG with  $V(K) = VB$  and  $E(K) = EW \cup EG$ , then  $K_{3,3} \not\rightarrow K$  iff  $K$  does not have a gray  $K_3$  or a gray  $K_{3,3}$ .

- If  $G$  has  $m$  edges (loops count 1/2), then

$$f_K(p) = \frac{1-2p}{2n} + p - \frac{pm}{n^2}.$$

## Theorem (Marchant-Thomason, 2010)

If  $p < 1/124$ , there is such a  $K$  for which

$$f_K(p) < \frac{p(1-p)}{1+p} = \gamma_{\text{Forb}(K_{3,3})}(p).$$

# Zarankiewicz constructions

W.G. Brown constructed dense  $K_{3,3}$ -free graphs.

## Lemma

If  $K$  is a CRG with  $V(K) = VB$  and  $E(K) = EW \cup EG$ , then  $K_{3,3} \not\rightarrow K$  iff  $K$  does not have a gray  $K_3$  or a gray  $K_{3,3}$ .

- If  $G$  has  $m$  edges (loops count 1/2), then

$$f_K(p) = \frac{1-2p}{2n} + p - \frac{pm}{n^2}.$$

## Theorem (Marchant-Thomason, 2010)

If  $p < 1/124$ , there is such a  $K$  for which

$$f_K(p) < \frac{p(1-p)}{1+p} = \gamma_{\text{Forb}(K_{3,3})}(p).$$

Something similar works for  $K_{2,t}$  (constructions due to Z. Füredi).

# What's next?

We have most of the tools we need in order to compute an edit distance function.



# What's next?

We have most of the tools we need in order to compute an edit distance function.

## *Part III:*

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