On the edit distance function of the random graph

Ryan Martin

Iowa State University

AMS Fall Central Sectional Meeting (#1171)
Omaha, NE (Virtual)

Joint work with
Alex W.N. Riasanovsky, Karlsruhe Institute of Technology

Cliques

Definition

If G and G' are graphs on the same labeled set of n vertices, then

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

(Classical) Question

Among all n-vertex graphs G, what is

$$\max \left\{ \mathrm{dist}(G, G') : G' \not\supseteq K_{p+1}, |V(G')| = n \right\}?$$

Cliques

Definition

If G and G' are graphs on the same labeled set of n vertices, then

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

(Classical) Question

Among all n-vertex graphs G, what is

$$\max\left\{\mathrm{dist}(G, G') : G' \not\supseteq K_{p+1}, |V(G')| = n\right\}?$$

Theorem (Turán, 1941)

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

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

Cliques

(Classical) Question

Among all n-vertex graphs G, what is

$$\max\left\{\mathrm{dist}(G,G') : G' \not\supseteq K_{p+1}, |V(G')| = n\right\}?$$

Theorem (Turán, 1941)

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

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

Answer

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

Extremal Edit Distance

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

Example: $\mathrm{Forb}(C_5)$, the property of having no induced copy of C_5.

Extremal Edit Distance

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

Example: $\mathrm{Forb}(C_5)$, the property of having no induced copy of C_5.

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 $\mathcal{H}$.

Generalizing the edit distance

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.

Generalizing the edit distance

Problem

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

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

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

The Edit Distance Function

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

- 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.
- $\mathrm{ed}_{\mathrm{Forb}(H)}(p) \leq \min\left\{\frac{p}{\chi(H)-1}, \frac{1-p}{\chi(\overline{H})-1}\right\}$,
  if neither H nor $\overline{H}$ is a clique.

The Edit Distance Function

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

- Continuous and concave down.
- $\mathrm{ed}_{\mathrm{Forb}(H)}(p) \leq \min\left\{\frac{p}{\chi(H)-1}, \frac{1-p}{\chi(\overline{H})-1}\right\}$,
  if neither H nor $\overline{H}$ is a clique.

Theorem (M.-Riasanovsky, 2021+)

Let $\mathcal{H} = \mathrm{Forb}(G(n_0, p_0))$. *Let* $\chi_0 = \frac{n_0}{2\log_{1/(1-p_0)} n_0}$ *and* $\overline{\chi_0} = \frac{n_0}{2\log_{1/p_0} n_0}$. *Let* $\varphi \approx 1.618$ *be the golden ratio. Then, a.a.s. as* $n \to \infty$,

$$\mathrm{ed}_{\mathcal{H}}(p) = (1+o(1))\min\left\{\frac{p}{\chi_0}, \frac{1-p}{\overline{\chi_0}}\right\}$$

- *for* $p_0 \in [1-1/\varphi, 1/\varphi] \approx [0.382, 0.618]$ *and all* $p \in [0,1]$, *or*
- *for* $p_0 \in (0,1)$ *and all* $p \in [1/3, 2/3]$.

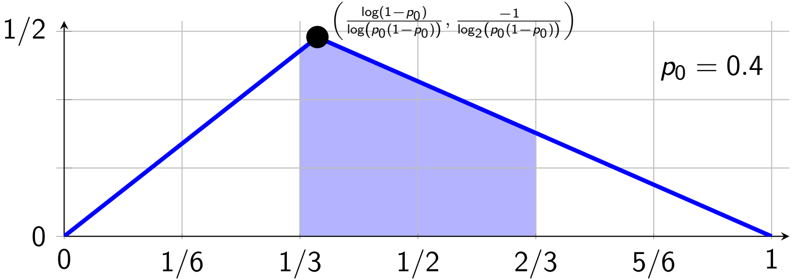
The Edit Distance Function

Theorem (M.-Riasanovsky, 2021+)

Let $\mathcal{H} = \mathrm{Forb}(G(n_0, p_0))$. *Then, a.a.s. as* $n \to \infty$,

$$\mathrm{ed}_{\mathcal{H}}(p) = (1 + o(1))\frac{2\log n_0}{n_0}\min\left\{\frac{p}{-\log(1-p_0)}, \frac{1-p}{-\log p_0}\right\}$$

- *for* $p_0 \in [1 - 1/\varphi, 1/\varphi] \approx [0.382, 0.618]$ *and all* $p \in [0, 1]$.



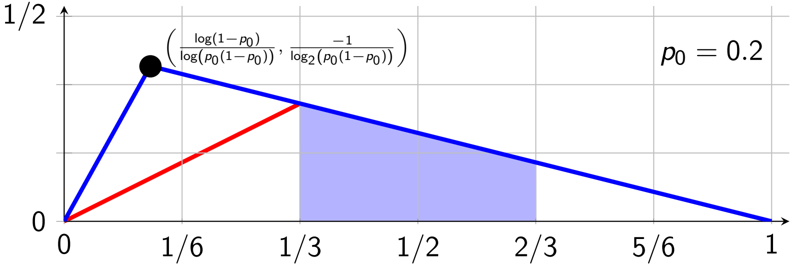
The Edit Distance Function

Theorem (M.-Riasanovsky, 2021+)

Let $\mathcal{H} = \mathrm{Forb}(G(n_0, p_0))$. *Then, a.a.s. as* $n \to \infty$,

$$\mathrm{ed}_{\mathcal{H}}(p) = (1 + o(1)) \frac{2 \log n_0}{n_0} \min \left\{ \frac{p}{-\log(1 - p_0)}, \frac{1 - p}{-\log p_0} \right\}$$

- *for* $p_0 \in (0, 1)$ *and all* $p \in [1/3, 2/3]$.



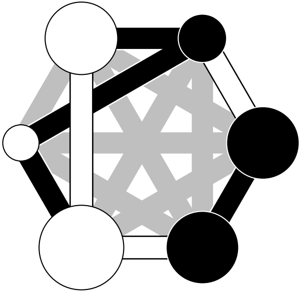
Colored Regularity Graphs

Definition

A COLORED REGULARITY GRAPH (CRG), $K = (V, E)$, is a complete graph such that

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

This is a recipe for editing.



Colored Regularity Graphs

Definition

COLORED REGULARITY GRAPH (CRG)

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

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

For each $\mathcal{H}$, there is a family of CRGs $\mathcal{K}(\mathcal{H})$ such that

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

where $g_K(p) = \min\{\mathrm{x}^T \mathrm{M} \mathrm{x} : \mathrm{x}^T \mathbf{1} = 1, \mathrm{x} \geq 0\}$ *for a matrix* $\mathrm{M} = \mathrm{M}_K(p)$.

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

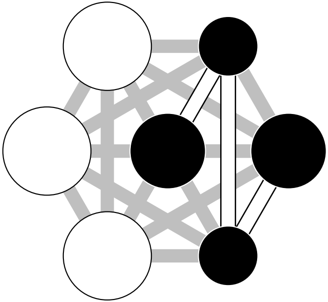
Colored Regularity Graphs

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

For each $\mathcal{H}$, there is a family of CRGs $\mathcal{K}(\mathcal{H})$ such that

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

where $g_K(p) = \min\{\mathrm{x}^T \mathrm{M} \mathrm{x} : \mathrm{x}^T \mathbf{1} = 1, \mathrm{x} \geq 0\}$ *for a matrix* $\mathrm{M} = \mathrm{M}_K(p)$.



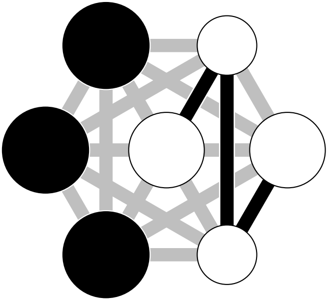
Colored Regularity Graphs

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

For each $\mathcal{H}$, there is a family of CRGs $\mathcal{K}(\mathcal{H})$ such that

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

where $g_K(p) = \min\{\mathrm{x}^T \mathrm{M} \mathrm{x} : \mathrm{x}^T \mathbf{1} = 1, \mathrm{x} \geq 0\}$ *for a matrix* $\mathrm{M} = \mathrm{M}_K(p)$.



Structure of p-cores

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

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

Structure of p-cores

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

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

This allows us to categorize p-core CRGs:

Theorem (Marchant-Thomason, 2010)

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

(a) $p \leq 1/2 \implies \mathrm{EB} = \emptyset$ *and no* EW *is incident to* VW.

(b) $p \geq 1/2 \implies \mathrm{EW} = \emptyset$ *and no* EB *is incident to* VB.

Structure of p-cores

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

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

This allows us to categorize p-core CRGs:

Theorem (Marchant-Thomason, 2010)

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

(a) $p \leq 1/2 \implies \mathrm{EB} = \emptyset$ *and no* EW *is incident to* VW.

(b) $p \geq 1/2 \implies \mathrm{EW} = \emptyset$ *and no* EB *is incident to* VB.

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

This is why we can compute $\mathrm{ed}_{\mathcal{H}}(1/2)$.

Underlying graphs

Definition

For a CRG, the UNDERLYING GRAPH is the graph formed by the non-gray edges.

- If $p < 1/2$, these are white edges (EW induced on VB).
- If $p > 1/2$, these are black edges (EB induced on VW).

Underlying graphs

Definition

For a CRG, the UNDERLYING GRAPH is the graph formed by the non-gray edges.

- If $p < 1/2$, these are white edges (EW induced on VB).
- If $p > 1/2$, these are black edges (EB induced on VW).

Definition

A graph G is p-PROHIBITED if the underlying graph of no p-core CRG has G as an induced subgraph.

Underlying graphs

Definition

A graph G is *p*-PROHIBITED if the underlying graph of no p-core CRG has G as an induced subgraph.

Lemma

If G is nonempty and $\lambda \leq -1$ is the minimum eigenvalue of the adjacency matrix of G, then G is p-prohibited if

$$p \in \left[\frac{1}{1-\lambda}, 1 - \frac{1}{1-\lambda}\right].$$

Underlying graphs

Lemma

If G is nonempty and $\lambda \leq -1$ is the minimum eigenvalue of the adjacency matrix of G, then G is p-prohibited if

$$p \in \left[\frac{1}{1-\lambda}, 1 - \frac{1}{1-\lambda}\right].$$

Fact

The minimum eigenvalue for P_3 is $-\sqrt{2}$.

$$\left[\frac{1}{1-(-\sqrt{2})}, 1 - \frac{1}{1-(-\sqrt{2})}\right] \approx [0.414, 0.586]$$

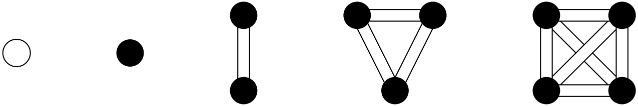
Thus, in this range every component of the underlying graph is a clique!

Dalmatian CRGs

Because of symmetry, we usually focus our attention to the case $p \leq 1/2$.

Definition

A COMPONENT of a CRG is a component of the underlying graph.
A component of a CRG that is a clique is a DALMATIAN CRG.

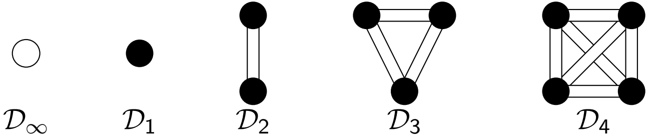


$\mathcal{D}_\infty$ $\mathcal{D}_1$ $\mathcal{D}_2$ $\mathcal{D}_3$ $\mathcal{D}_4$

Lemma (Dalmatian Interval)

*If $p \in \left[\sqrt{2}-1, 2-\sqrt{2}\right] \approx [0.414, 0.586]$,
then every component of a p-core CRG is dalmatian.*

Dalmatian CRGs



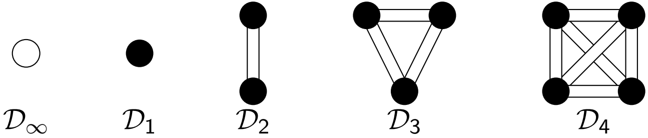
Lemma (Dalmatian Interval)

If $p \in \left[\sqrt{2}-1, 2-\sqrt{2}\right] \approx [0.414, 0.586]$,
then every component of a p*-core CRG is dalmatian.*

Lemma (Improved Dalmatian Interval)

If $p \in [1-1/\varphi, 1/\varphi] \approx [0.382, 0.618]$,
then every component of a p*-core CRG is dalmatian.*

Dalmatian CRGs



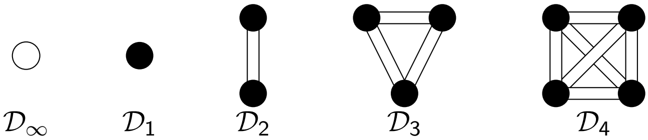
Lemma (Improved Dalmatian Interval)

If $p \in [1 - 1/\varphi, 1/\varphi] \approx [0.382, 0.618]$,
then every component of a p-core CRG is dalmatian.

Proof.

- C_4 is p-prohibited for $p \in [1/3, 2/3]$.
- P_4 is p-prohibited for $p \in [1 - 1/\varphi, 1/\varphi]$.
- A connected $\{C_4, P_4\}$-free graph has a dominant vertex.

Dalmatian CRGs



Lemma (Improved Dalmatian Interval)

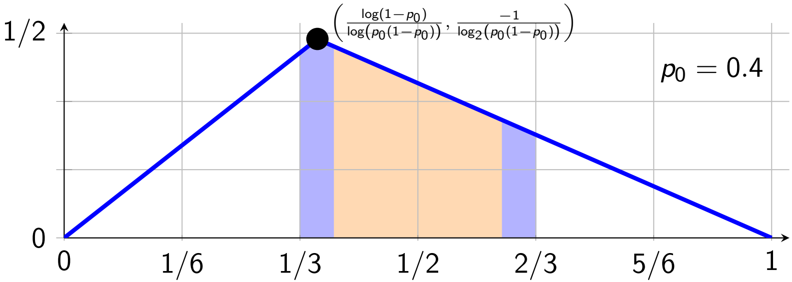
If $p \in [1 - 1/\varphi, 1/\varphi] \approx [0.382, 0.618]$,
then every component of a p-core CRG is dalmatian.

Proof.

- C_4 is p-prohibited for $p \in [1/3, 2/3]$.
- P_4 is p-prohibited for $p \in [1 - 1/\varphi, 1/\varphi]$.
- A connected $\{C_4, P_4\}$-free graph has a dominant vertex.

If $p \in [1/3, 2/3]$, then any component with a dominant vertex is a clique. Hence, in the in smallest interval, $[1 - 1/\varphi, 1/\varphi]$, all components are cliques. □

Two different roles for φ



- $[1 - 1/\varphi, 1/\varphi]$
- $[1/3, 2/3]$

Although we know the structure of CRGs in the orange range, we are able to get the full range $p \in [1/3, 2/3]$ for the random graph $H = G(n_0, p_0)$.

Bounded components

Lemma

Fix $p \in (1/3, 2/3)$ *and* $\varepsilon \in (0, 1)$. *There exists a positive integer* $B = B(p, \varepsilon)$ *such that the following holds: For all CRGs* K, *there exists a* p*-core sub-CRG* K' *whose components have order at most* B *and* $g_{K'}(p) \leq (1 + \varepsilon) g_K(p)$.

Bounded components

Lemma

Fix $p \in (1/3, 2/3)$ and $\varepsilon \in (0,1)$. There exists a positive integer $B = B(p,\varepsilon)$ such that the following holds: For all CRGs K, there exists a p-core sub-CRG K' whose components have order at most B and $g_{K'}(p) \leq (1+\varepsilon)g_K(p)$.

Proof idea:

- Trim the CRG to get bounded degree, but barely-changed g function.
- The diameter is bounded because

$$\lambda(P_d) = -2\cos\left(\frac{\pi}{d+1}\right) \to -2.$$

- Bounded degree and bounded diameter yields bounded order.

Bounded components

Lemma

Fix $p \in (1/3, 2/3)$ and $\varepsilon \in (0,1)$. There exists a positive integer $B = B(p,\varepsilon)$ such that the following holds: For all CRGs K, there exists a p-core sub-CRG K' whose components have order at most B and $g_{K'}(p) \le (1+\varepsilon) g_K(p)$.

Proof idea:

- Trim the CRG to get bounded degree, but barely-changed g function.
- The diameter is bounded because

$$\lambda(P_d) = -2\cos\left(\frac{\pi}{d+1}\right) \to -2.$$

- Bounded degree and bounded diameter yields bounded order.

Trimming to bound the degree works for all p.
Bounded diameter only works for $p \in [1/3, 2/3]$.

Future work

The conjecture is still open for $p_0 \in (0, 1-1/\varphi) \cup (1/\varphi, 1)$:

Conjecture

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

$$\mathrm{ed}_{\mathcal{H}}(p) \sim \frac{2\log_2 n_0}{n_0} \min\left\{ \frac{p}{-\log_2(1-p_0)}, \frac{1-p}{-\log_2 p_0} \right\}.$$

with probability $\to 1$ as $n_0 \to \infty$. Note:

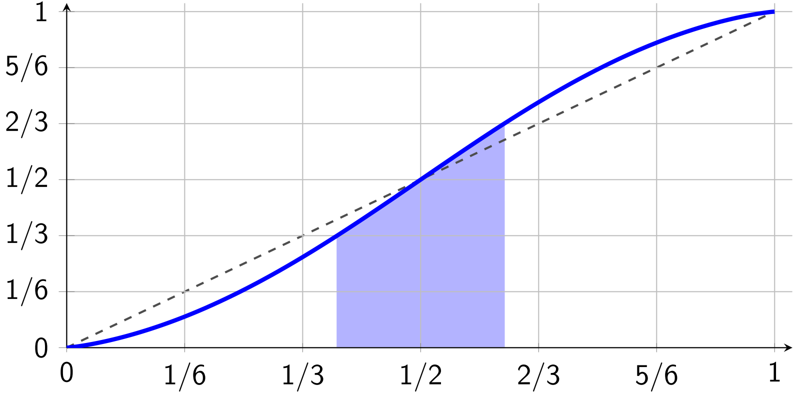
$$p^*_{\mathcal{H}} \sim \frac{\log(1-p_0)}{\log(p_0(1-p_0))}.$$

Counterintuitive because $\frac{\log(1-p_0)}{\log(p_0(1-p_0))} = p_0$ if and only if $p_0 \in \left\{0, \frac{1}{2}, 1\right\}$.

The main barrier is that we cannot bound the number of possibilities for a CRG component, which was essential for our proof. Perhaps we can circumvent that.

Future work

$$p^*_{\mathcal{H}} \sim \frac{\log(1-p_0)}{\log(p_0(1-p_0))}$$



Future work

- If $p \in [1 - 1/\varphi, 1/\varphi] = [2 - \varphi, \varphi - 1] \approx [0.382, 618]$, then every component of a p-core CRG is a clique (dalmatian).

  What do p-core CRGs look like in the next interval and how wide is it? (In progress)

- What are some other properties of the edit distance function, such as how few CRGs are necessary to define the whole function? (Submitted with Cox and McGinnis)

- The edit distance can be defined for hypergraphs, but the notoriously difficult hypergraph Turán problem is a special case ($p = 1$).
  - Is there anything interesting that can be said for hypergraphs?
  - Is there an interval of p for which hypergraph edit distance functions can be computed?
  - Is it possible that this theory might shed light on the hypergraph Turán problem in some open cases?