PHOEG Helps Obtaining Extremal Graphs

Hadrien Mélot

Joint work with Gauvain Devillez and Pierre Hauweele

Service d'Algorithmique, Département d'Informatique

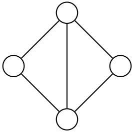
Faculté des Sciences, Université de Mons

UMONS

GGTW, Ghent, August 18, 2017

Introduction

We consider simple undirected graphs.



For a graph $G = (V, E)$,

- its order $|V|$ is denoted by n;
- its size $|E|$ is denoted by m.

Introduction

- Context: Computer-assisted Discovery in Extremal Graph Theory
- Several existing systems: Graph, Graffiti, AutoGraphiX, GraPHedron, ...
  - exploit different ideas to help graph theorists
- Objectives of this talk:
  - presentation of PHOEG, a successor of GraPHedron
  - use of an illustrative problem (eccentric connectivity index, ECI)
- Remark: work under progress
  - PHOEG is currently a prototype
  - the problem about ECI is not fully solved

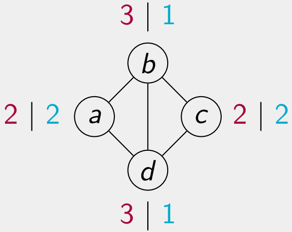
Eccentric Connectivity Index

Definition

The Eccentric Connectivity Index (ECI) of a graph G, denoted by $\xi^c(G)$, is

$$\xi^c(G) = \sum_{v \in V} d(v)\epsilon(v).$$

Example



$$\xi^c(G) = 2 \times (4 + 3) = 14$$

Upper bound on ξ^c for connected graphs with fixed size

Problem

Among connected graphs of order n and size m, what is the maximum possible value for ξ^c?

(To avoid infinite eccentricities, we restrict the problem to connected graph)

Graphs $E_{n,m}$

We define $E_{n,m}$ as follows :

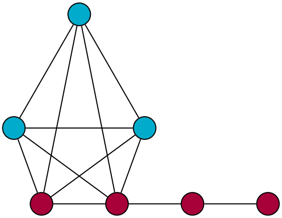
$$n = 7, m = 14$$

Graphs $E_{n,m}$

We define $E_{n,m}$ as follows :

- The biggest possible clique without disconnecting the graph, leaving a path with the remaining vertices.

$n = 7, m = 14$

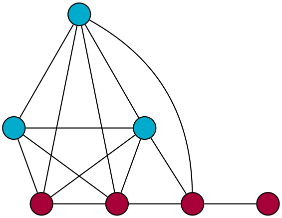


Graphs $E_{n,m}$

We define $E_{n,m}$ as follows :

- The biggest possible clique without disconnecting the graph, leaving a path with the remaining vertices.
- Add remaining edges between vertices of the clique and the first vertex of the path.

$n = 7, m = 14$

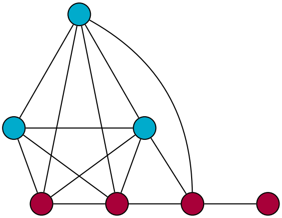


Graphs $E_{n,m}$

We define $E_{n,m}$ as follows :

- The biggest possible clique without disconnecting the graph, leaving a path with the remaining vertices.
- Add remaining edges between vertices of the clique and the first vertex of the path.

$n = 7, m = 14$



This graph is unique for given n and m.

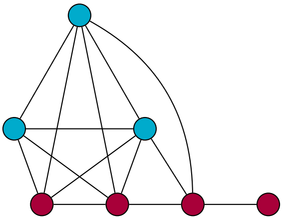
Graphs $E_{n,m}$

For positive integers n and m with $n-1 \leq m \leq \binom{n}{2}$, let define

$$d_{n,m} = \left\lfloor \frac{2n+1-\sqrt{17+8(m-n)}}{2} \right\rfloor .$$

Remarks :

- In the following, we simply use d for $d_{n,m}$;
- d is decreasing when m increase (and n fixed);
- For $E_{n,m}$, d is its diameter.



Conjecture of Zhang, Liu and Zhou

Conjecture (Zhang, Liu and Zhou, 2014)

Let G be a connected graph of order n and size m such that $d \geq 3$. Then,

$$\xi^c(G) \leq \xi^c(E_{n,m}),$$

with equality if and only if $G \simeq E_{n,m}$.

- The authors prove that the conjecture is true when $m = n-1, n, \ldots, n+4$ (if n is large enough).
- It exists a "proof" published in a journal of University of Isfahan (Iran, 2014) but that is obviously wrong.

Conjecture of Zhang, Liu and Zhou

Conjecture (Zhang, Liu and Zhou, 2014)

Let G be a connected graph of order n and size m such that $d \geq 3$. Then, $\xi^c(G) \leq \xi^c(E_{n,m})$, with equality iff $G \simeq E_{n,m}$.

This conjecture leads to several questions:

- Is the conjecture true?
- If yes, how to prove it?
- If no, how to improve or correct it?
- What about graphs such that $d \leq 2$?

How the computer can help?

In the following, we will show how PHOEG can help study all of the above questions and raise new ones.

P $\quad$ H$_{\text{elps}}$ $\quad$ O$_{\text{btaining}}$ E$_{\text{xtremal}}$ G$_{\text{raphs}}$

Overview of PHOEG

- Former system (GraPHedron): graphs and invariant's values written sequentially in files;
- PHOEG uses a PostgreSQL DB with more than 12 millions of non-isomorphic graphs (up to order 10) and tables of corresponding invariants' values;
- Invariant's values are computed once (useful for NP-hard invariants);
- Each graph has its unique signature used as primary key (canonical form, thanks to Nauty)
- This allows complex (and fast) queries on graphs.

Database query – Points and multiplicities

```
SELECT P.val AS eci, num_edges.val AS m,
  COUNT(*) AS mult
FROM eci P
  JOIN num_vertices USING(signature)
  JOIN num_edges USING(signature)
WHERE num_vertices.val = 7
GROUP BY m, eci;
```

```
eci | m  | mult
----+----+------
 47 |  8 |    5
 46 |  8 |    3
 40 |  8 |    3
 32 |  7 |    3
 48 | 12 |   55
 48 | 18 |    1
 61 | 14 |    4
 59 | 13 |    1
 48 | 11 |   17
 43 |  9 |   14
 47 |  6 |    1
 64 | 10 |    1
 59 | 11 |    1
 45 |  9 |    7
 38 |  6 |    2
     [...]
```

Database query – Polytope

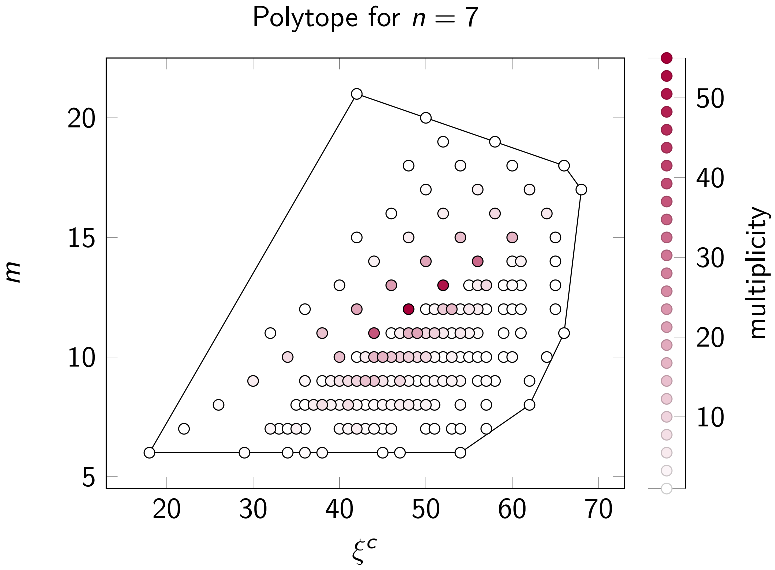
Main principle:

- view graphs as points in the space of invariants;
- compute the convex hull of these points (for small values of n).

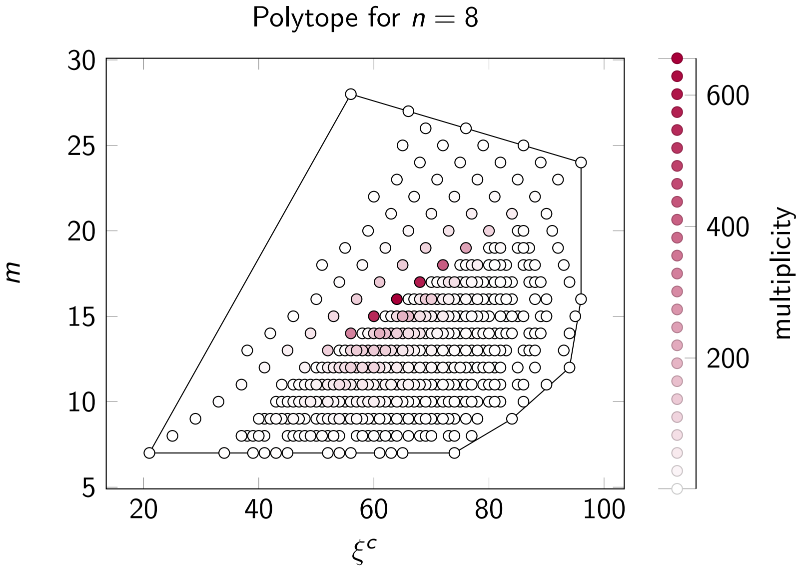
```
SELECT ST_AsText(ST_ConvexHull(ST_Collect(ST_Point(eci, m))))
FROM poly;

                    st_astext
-------------------------------------------------------
 POLYGON((18 6,42 21,66 18,68 17,66 11,62 8,54 6,18 6))
```

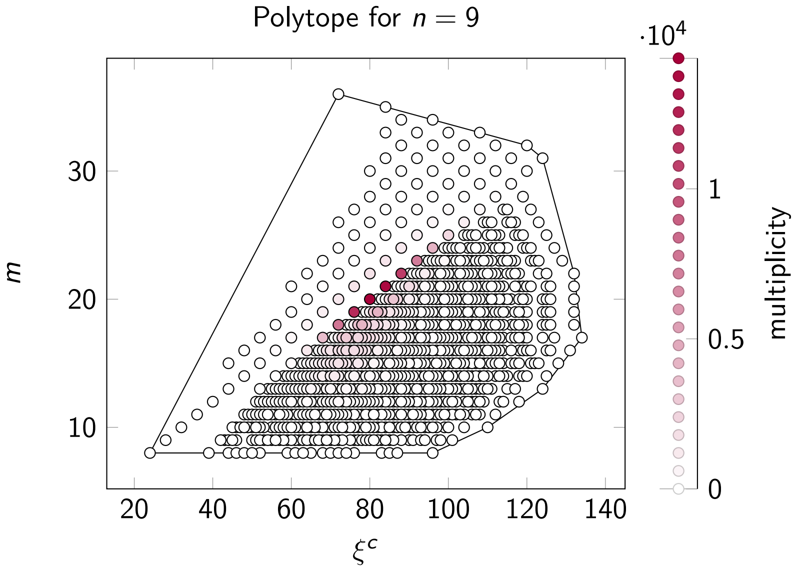
Exploring ξ^c with PHOEG: polytopes



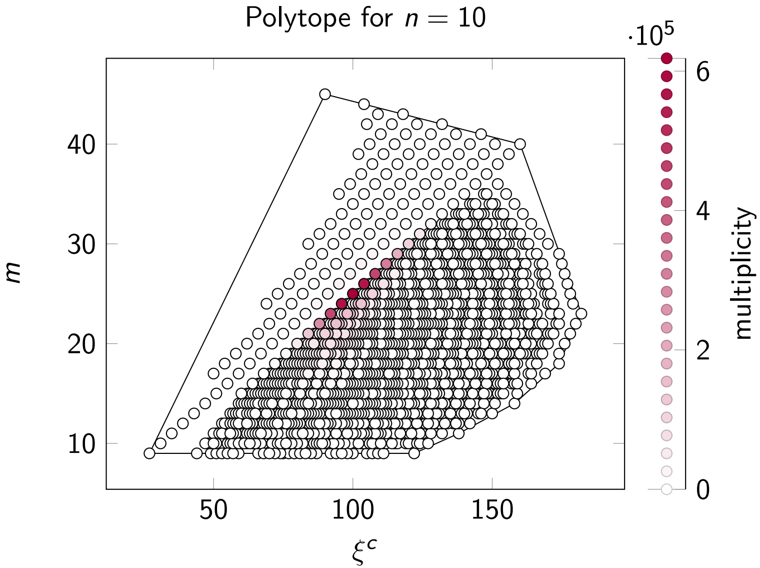
Exploring ξ^c with PHOEG: polytopes



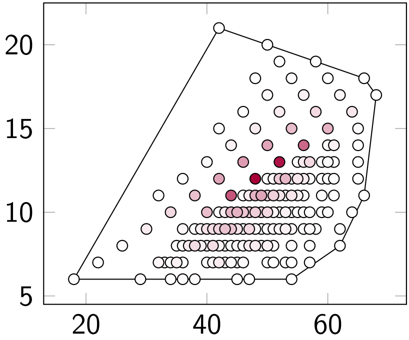
Exploring ξ^c with PHOEG: polytopes



Exploring ξ^c with PHOEG: polytopes



Observations and questions



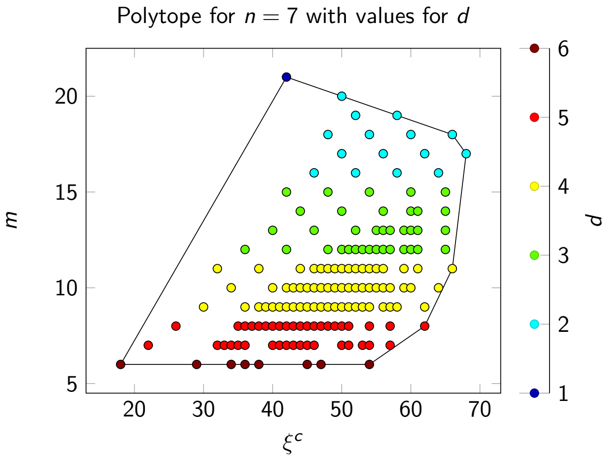
- Is the conjecture of Zhang, Liu and Zhou true when $d \geq 3$?
- Upper bound when $d \leq 2$?
- How to explain the grid?

Database query – adding other information

```
SELECT num_edges.val AS m,
   p.val AS eci, d.val AS d,
   diam.val AS diam
FROM eci p
  JOIN num_vertices USING(signature)
  JOIN num_edges USING(signature)
  JOIN d USING(signature)
  JOIN diam USING(signature)
WHERE num_vertices.val = 7
ORDER BY diam, d, m, eci;
```

```
m  | eci | d | diam
---+-----+---+------
21 |  42 | 1 |     1
16 |  46 | 2 |     2
16 |  52 | 2 |     2
16 |  52 | 2 |     2
16 |  52 | 2 |     2
16 |  52 | 2 |     2
16 |  52 | 2 |     2
16 |  58 | 2 |     2
16 |  58 | 2 |     2
16 |  58 | 2 |     2
16 |  58 | 2 |     2
16 |  58 | 2 |     2
16 |  58 | 2 |     2
16 |  58 | 2 |     2
16 |  58 | 2 |     2
          [...]
```

Coloring points with values of d



Is the conjecture true for $d \geq 3$?

Database query – Extremal graphs

```
WITH tmp AS (
  SELECT n.val AS n, m.val AS m,
    P.signature, P.val AS eci, d.val AS d
    rank() OVER (
      PARTITION BY n.val, m.val
      ORDER BY P.val DESC
    ) AS pos
  FROM num_vertices n
  JOIN num_edges m USING(signature)
  JOIN d USING(signature)
  JOIN eci P USING(signature)
  WHERE n.val = 7
)
SELECT signature AS sig, n, m, eci, d
FROM tmp
WHERE pos = 1 AND d >= 3
ORDER BY n, m, d, eci;
```

```
 sig   | n | m  | eci | d
-------+---+----+-----+----
F@IQO  | 7 |  6 |  54 | 6
F@‘J_  | 7 |  7 |  57 | 5
FgCXW  | 7 |  8 |  62 | 5
FWCYw  | 7 |  9 |  62 | 4
FgCxw  | 7 | 10 |  64 | 4
F‘Kyw  | 7 | 11 |  66 | 4
F‘Kzw  | 7 | 12 |  65 | 3
F‘Lzw  | 7 | 13 |  65 | 3
F‘\zw  | 7 | 14 |  65 | 3
FJ]|w  | 7 | 15 |  65 | 3
FJ\|w  | 7 | 15 |  65 | 3
```

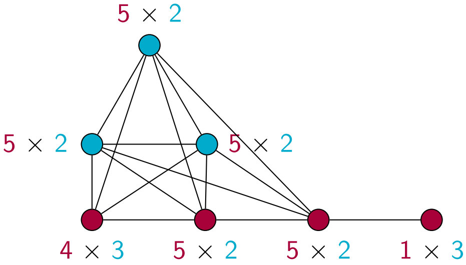
Database query – Extremal graphs

```
WITH tmp AS (
  SELECT n.val AS n, m.val AS m,
    P.signature, P.val AS eci, d.val AS d
    rank() OVER (
      PARTITION BY n.val, m.val
      ORDER BY P.val DESC
    ) AS pos
  FROM num_vertices n
  JOIN num_edges m USING(signature)
  JOIN d USING(signature)
  JOIN eci P USING(signature)
  WHERE n.val = 7
)
SELECT signature AS sig, n, m, eci, d
FROM tmp
WHERE pos = 1 AND d >= 3
ORDER BY n, m, d, eci;
```

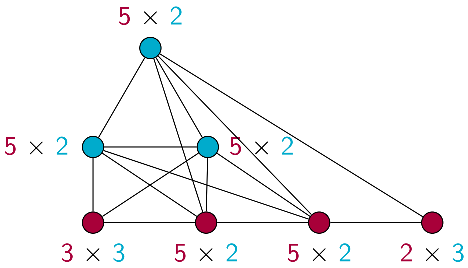
```
  sig  | n | m  | eci | d
-------+---+----+-----+----
 F@IQO | 7 |  6 |  54 | 6
 F@‘J_ | 7 |  7 |  57 | 5
 FgCXW | 7 |  8 |  62 | 5
 FWCYw | 7 |  9 |  62 | 4
 FgCxw | 7 | 10 |  64 | 4
 F‘Kyw | 7 | 11 |  66 | 4
 F‘Kzw | 7 | 12 |  65 | 3
 F‘Lzw | 7 | 13 |  65 | 3
 F‘\zw | 7 | 14 |  65 | 3
 FJ]|w | 7 | 15 |  65 | 3
 FJ\|w | 7 | 15 |  65 | 3
```

Counter-example to the conjecture: no unicity for extremal graphs

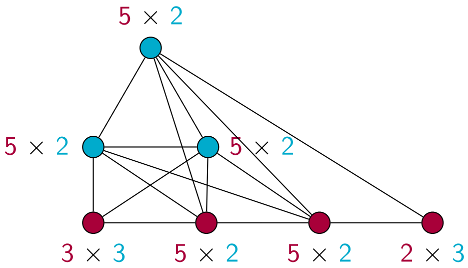
Counter-example ($n = 7$ and $m = 15$)



Counter-example ($n = 7$ and $m = 15$)

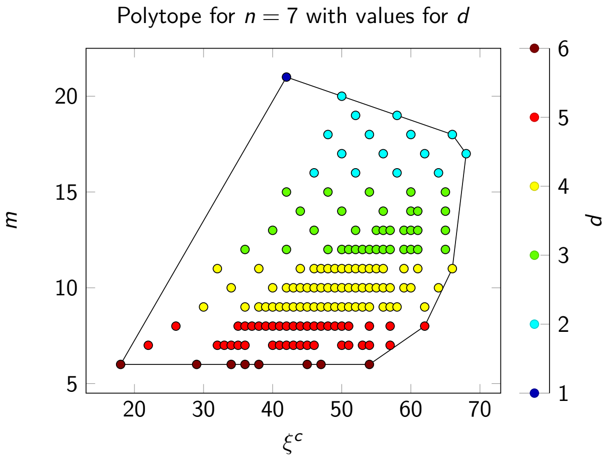


Counter-example ($n = 7$ and $m = 15$)



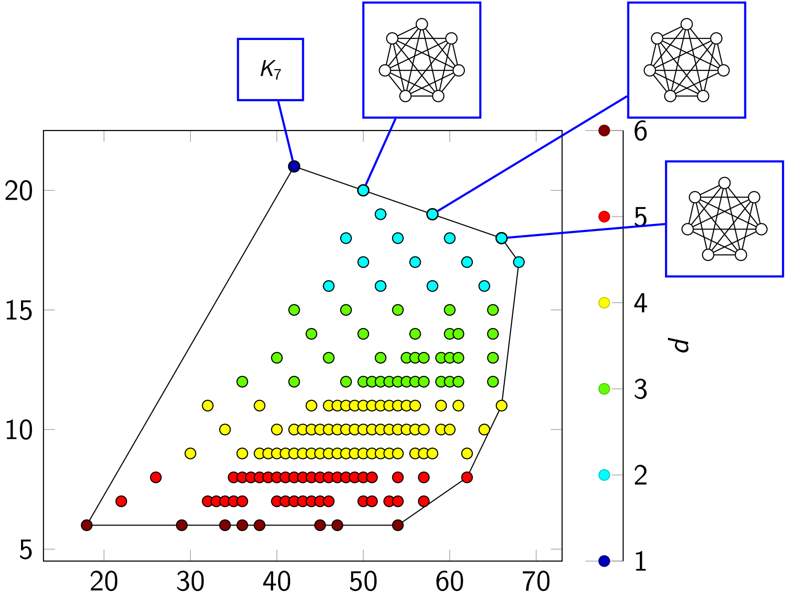
It is possible to construct counter-examples for any values of $n \geq 6$ (with $d = 3$).

Coloring points with values of d



Upper bound when $d \leq 2$?

Upper facet of the polytope ($n = 7$)



A new upper bound tight when $d \leq 2$

Theorem

Let G be a graph of order n and size m. Then,

$$\xi^c(G) \leq n(n-1)(n-2) - 2m(n-3),$$

with equality if and only if G is the complement of a matching.

Note that the bound is valid for all graphs but can be tight only if

$$m \geq \binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor,$$

(and thus $d \leq 2$).

PHOEG – Transproof

Up to this point, we have

- a tight upper bound when $d \leq 2$;
- and counter-examples for the unicity if $d = 3$.

It is also possible to have an explanation of the grid using PHOEG (skipped).

Is the conjecture true if $d \geq 4$? (actually, we believe it is).

If yes, is PHOEG able to help for a proof?

PHOEG – Transproof

Up to this point, we have

- a tight upper bound when $d \leq 2$;
- and counter-examples for the unicity if $d = 3$.

It is also possible to have an explanation of the grid using PHOEG (skipped).

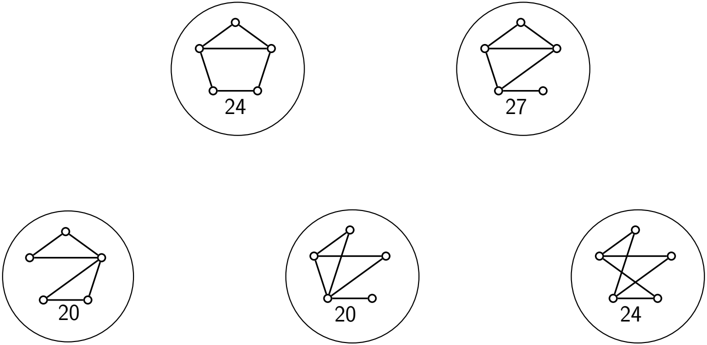
Is the conjecture true if $d \geq 4$? (actually, we believe it is).

If yes, is PHOEG able to help for a proof?

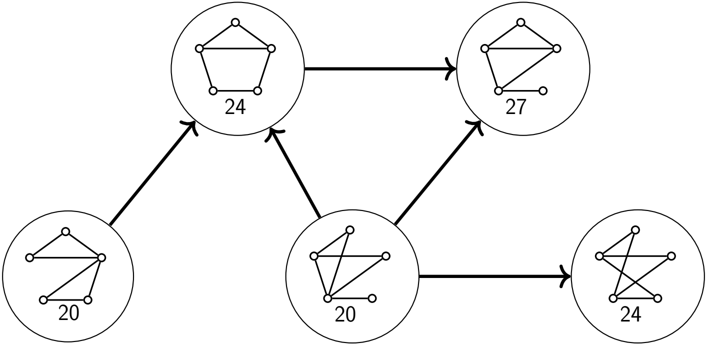
This is the purpose of the *Transproof* module:

- using graph transformations is a common proof technique;
- not always easy to find "good" transformations.

Metagraph of rotations that increase ξ^c ($n = 5$ and $m = 6$)



Metagraph of rotations that increase ξ^c ($n = 5$ and $m = 6$)



Applying only one rotation is thus not sufficient to have a proof.

Metagraph of transformations

- Idea : find an "improving path" from any graph to an extremal graph
- Metagraph stored in a graph DB (*Neo4j*)
  - for all graphs up to order 9
  - using 8 simple transformations
  - $\rightarrow$ for $n = 9$, it makes a metagraph of order 274 668 and size 380 814 904
- Useful to reject / find good transformations while searching for a proof

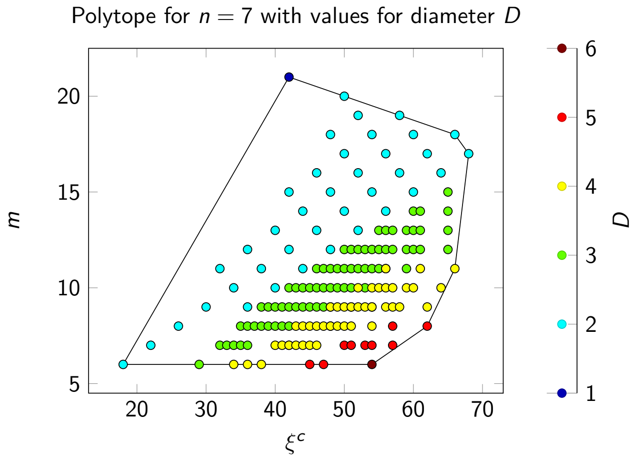
Concluding remarks

- Not only extremal graphs are useful to study extremal properties of an invariant
- Exact approach limited to small graphs ($n \leq 10$)
- However, dealing with small graphs has already shown to be very useful and led to numerous results (AutoGraphiX, GraPHedron)
- PHOEG is intended to be an ecosystem of useful modules (e.g., Forbidden Subgraph Characterization)

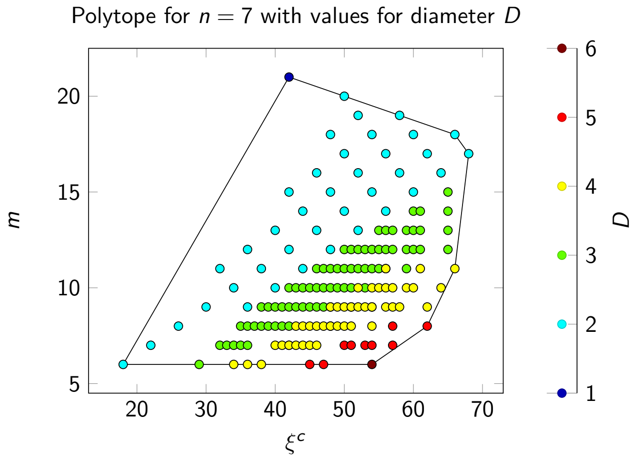
Perspectives

- Invariants' DB allows a form of dynamic programming
- Create a simple interface for queries
- Allow easy visualization and manipulation of outputs (GUI, PDF, etc.)
- Simplify the definitions of transformations
- Suggest automatically (a short list of) transformations

Coloring points with highest diameter

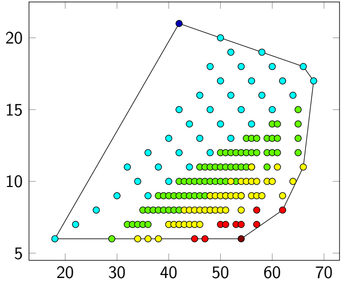


Coloring points with highest diameter



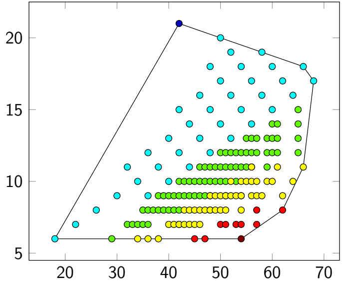
Can the diameter explain the blue grid? Actually, yes!

Understanding the grid of blue points



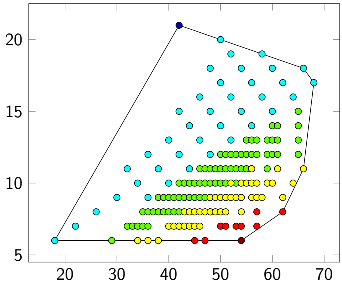
- Suppose $D(G) = 2$ (light blue points)
- For each vertex v, since $D(G) = 2$, either $\epsilon(v) = 1$ or $\epsilon(v) = 2$

Understanding the grid of blue points



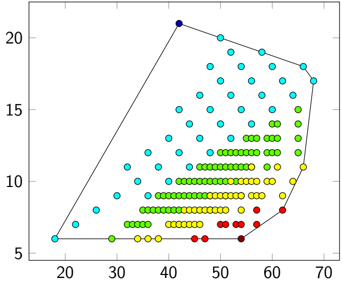
- Suppose $D(G) = 2$ (light blue points)
- For each vertex v, since $D(G) = 2$, either $\epsilon(v) = 1$ or $\epsilon(v) = 2$
- If $\epsilon(v) = 1$, then v is dominant and $d(v) = n - 1$

Understanding the grid of blue points



- Suppose $D(G) = 2$ (light blue points)
- For each vertex v, since $D(G) = 2$, either $\epsilon(v) = 1$ or $\epsilon(v) = 2$
- If $\epsilon(v) = 1$, then v is dominant and $d(v) = n - 1$
- Let k be the number of dominant vertices of G

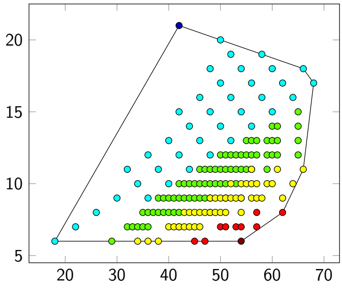
Understanding the grid of blue points



- Suppose $D(G) = 2$ (light blue points)
- For each vertex v, since $D(G) = 2$, either $\epsilon(v) = 1$ or $\epsilon(v) = 2$
- If $\epsilon(v) = 1$, then v is dominant and $d(v) = n - 1$
- Let k be the number of dominant vertices of G
- The sum of degrees of non dominant vertices is

$$2m - k(n-1)$$

Understanding the grid of blue points



- Suppose $D(G) = 2$ (light blue points)
- For each vertex v, since $D(G) = 2$, either $\epsilon(v) = 1$ or $\epsilon(v) = 2$
- If $\epsilon(v) = 1$, then v is dominant and $d(v) = n - 1$
- Let k be the number of dominant vertices of G
- The sum of degrees of non dominant vertices is

$$2m - k(n-1)$$

Thus,

$$\xi^c(G) = k(n-1) + 2(2m - k(n-1)) = 4m - k(n-1),$$

that is maximum if $k = 0$ and, moreover, explain the grid.