PHOEG Helps Obtaining Extremal Graphs

Pierre Hauweele

Joint work with Gauvain Devillez and Hadrien Mélot

Algorithms Lab, Computer Science Departement

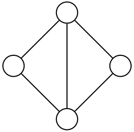
Faculty of Sciences, University of Mons

UMONS

CSD8, Mons, August 23, 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.

A graph invariant is a function on graphs that is constant on isomorphism classes.

Examples: order n, size m, chromatic number χ, maximum degree Δ, diameter D, planarity, ...

Extremal Graph Theory

Extremal Graph Theory aims to find bounds on a graph invariant under some constraints.
Generally, those constraints are of two types:

- restricting class of graphs (e.g., connected graphs, trees);
- fixing (and restricting) values of other invariants (e.g., size, maximum degree).

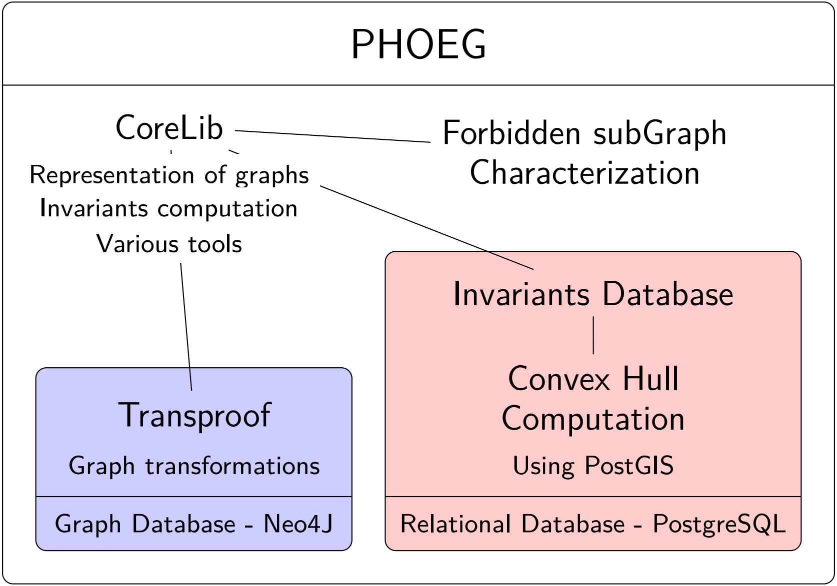
Results in Extremal Graph Theory mainly consists in

- giving bounds;
- characterizing graphs achieving these bounds.

Computer-assisted discovery

- 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 in progress
  - PHOEG is currently a prototype
  - the problem about ECI is not fully solved

Overview of PHOEG

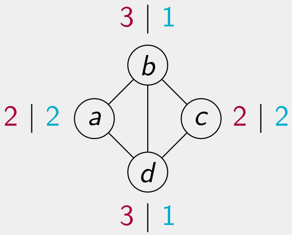


Eccentric Connectivity Index

Let v be a vertex of a graph G, recall that:

- degree $d(v)$ = number of adjacent vertices of v;
- eccentricity $\epsilon(v)$ = maximal distance between v and any other vertex.

Example



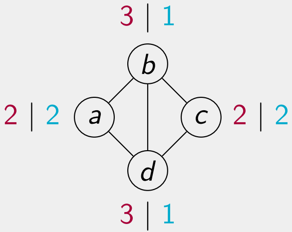
Eccentric Connectivity Index

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



$$\xi^c(G) = (2 \times 2 + 3 \times 1) \times 2 = 14$$

Eccentric Connectivity Index

History and motivation

- Sharma, Goswani and Madan introduced ξ^c in 1997 in Chemistry;
- Useful as a discriminating topological descriptor for Structure Properties and Structure Activity studies;
- Since 1997, more than 200 chemical papers about ξ^c: applications in drug design, prediction of anti-HIV activities, etc.
- However, the first mathematical paper with extremal properties on ξ^c was published only in 2010;
- Since 2010, about a dozen papers containing bounds on ξ^c.

Some Extremal Theory problem about ξ^c

Now, let's make extremal graph theory about ξ^c with the help of a computer.

First step: define a problem by choosing constraints.

Some Extremal Theory problem about ξ^c

Now, let's make extremal graph theory about ξ^c with the help of a computer.

First step: define a problem by choosing constraints.

Several papers containing bounds on ξ^c — using various invariants as constraints — have been published (since 2010).

Some Extremal Theory problem about ξ^c

Now, let's make extremal graph theory about ξ^c with the help of a computer.

First step: define a problem by choosing constraints.

Several papers containing bounds on ξ^c — using various invariants as constraints — have been published (since 2010).

Problem

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

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

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

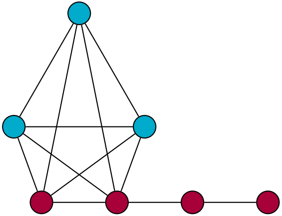
$n = 7, m = 14$

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

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

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

$n = 7, m = 14$

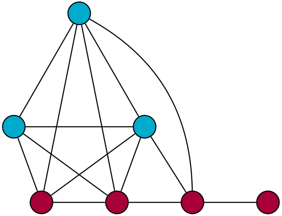


Upper bound on ξ^c for connected graphs with fixed size

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.

$n = 7, m = 14$

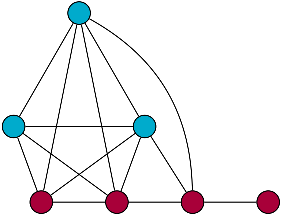


Upper bound on ξ^c for connected graphs with fixed size

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. We define $d_{n,m}$ as the diameter of $E_{n,m}$.

Conjecture of Zhang, Liu and Zhou

Conjecture (Zhang, Liu and Zhou, 2014)

Let G be a graph of order n and size m such that $d_{n,m} \geq 3$. 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).
- There 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 graph of order n and size m such that $d_{n,m} \geq 3$. Then,

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

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

- Is the conjecture true?
- If yes, how to prove it?
- If no, how to improve or correct it?
- What about graphs such that $d_{n,m} < 3$?

How can the computer help?

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

P Helps Obtaining Extremal Graphs

PHOEG — the database part

- Former system (GraPHedron): graphs and invariant's values written sequentially in files;
- PHOEG uses a PostgreSQL DB with tens of millions of non-isomorphic graphs and invariants' values;
- Invariant's values are computed once (useful for NP-hard invariants);

Database of the invariants

- Each graph has its unique signature used as primary key (canonical form, thanks to Nauty by Brendan McKay), $sig(C_5) = "DqK"$, $sig(K_3) = "Bw"$.
- 12 millions simple graphs up to order 10, 8 millions cubic graphs up to order 22.

Graphs
signature
A_
A?
B?
BG
Bw
BW
C`
C^
C~
C?
C@

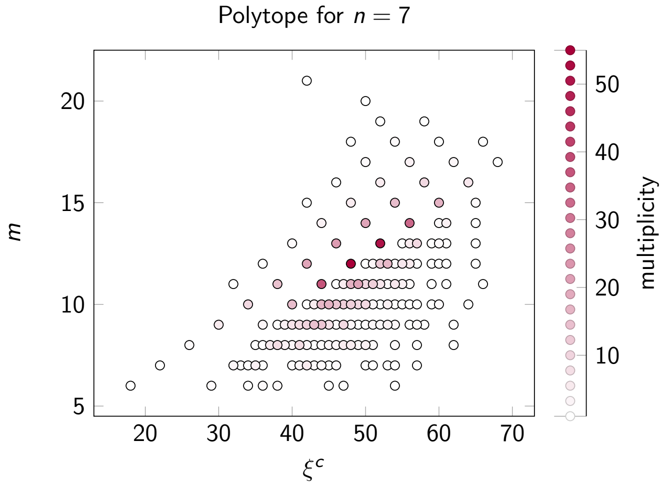
NumVertices	
signature	val
A_	2
A?	2
B?	3
BG	3
Bw	3
BW	3
C`	4
C^	4
C~	4
C?	4
C@	4

NumEdges	
signature	val
A_	1
A?	0
B?	0
BG	1
Bw	3
BW	2
C`	2
C^	5
C~	6
C?	0
C@	1

ECI	
signature	val
A_	2
BW	6
Bw	6
C^	14
C~	12
CF	9
CN	13
Cr	16
CR	14
D`[	25
D`{	20

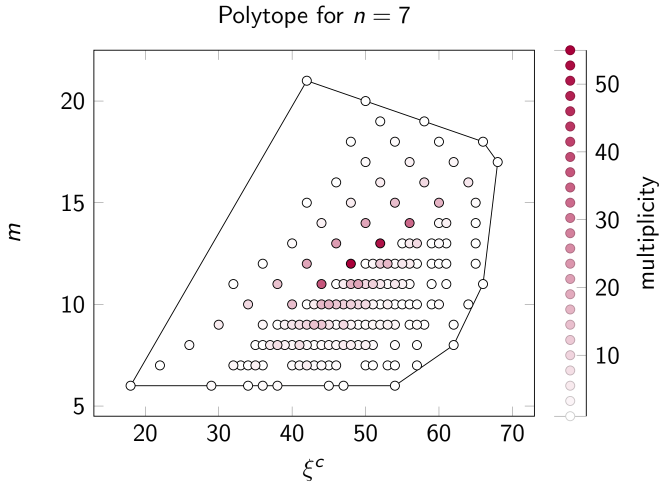
GraPHedron's main principle

- view graphs as points in the space of invariants;



GraPHedron's main principle

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



Database query – Points, multiplicities and polytope

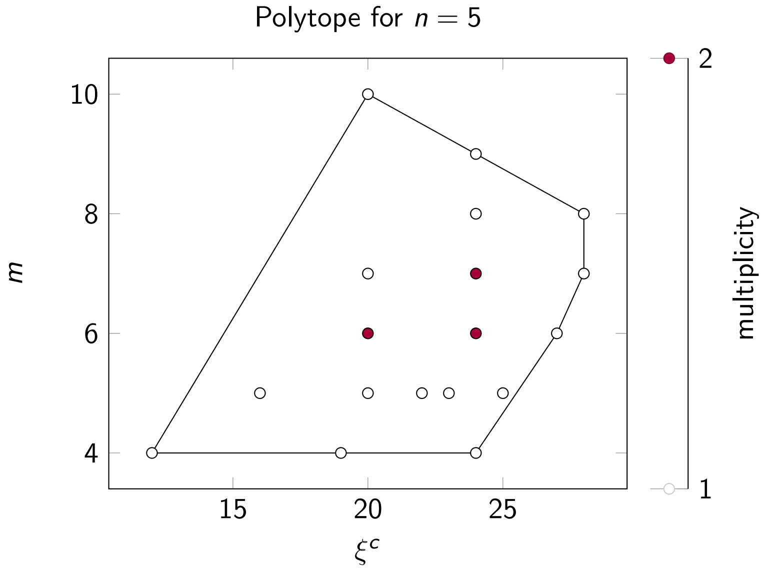
```
SELECT P.val AS eci, num_edges.val AS m,
  COUNT(*) AS mult
FROM eci P
  JOIN num_vertices USING(signature)
  JOIN num_edges USING(signature)
WHERE num_vertices.val = 7
GROUP BY m, eci;
```

```
 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
         [...]
```

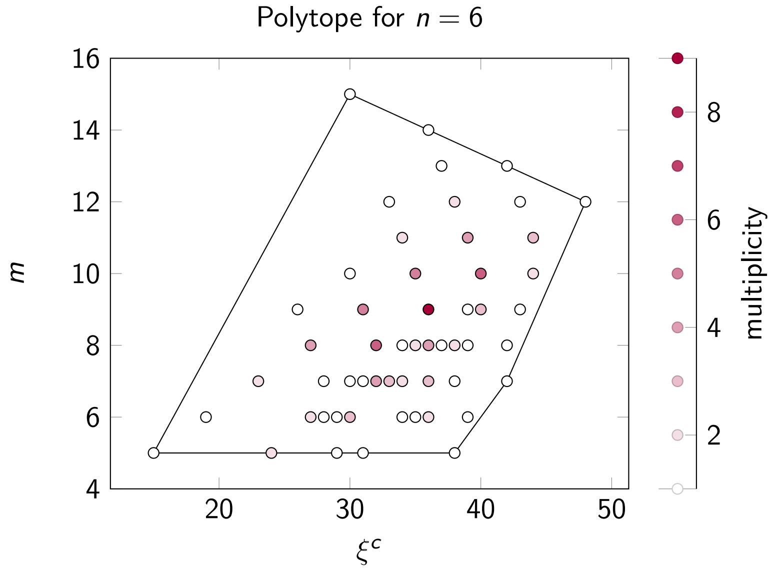
```
SELECT ST_AsText(ST_ConvexHull(
  ST_Collect(ST_Point(eci, m))))
FROM poly;
```

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

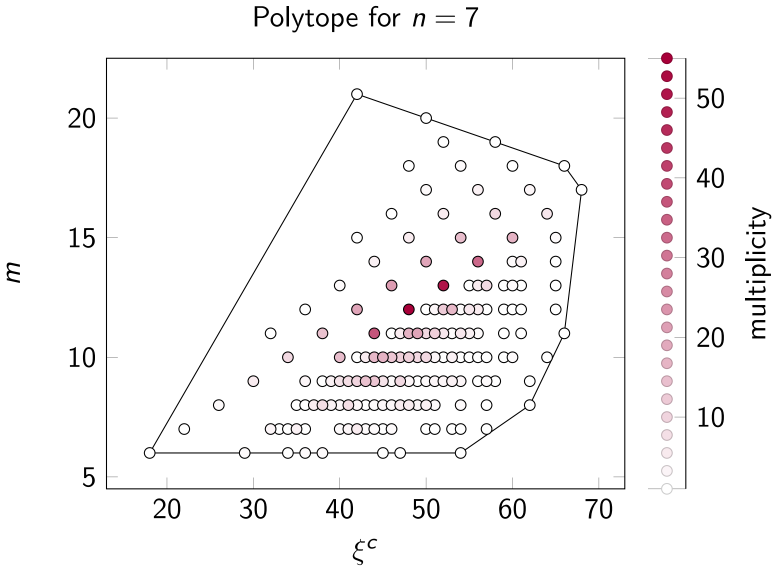
Exploring ξ^c with PHOEG: polytopes



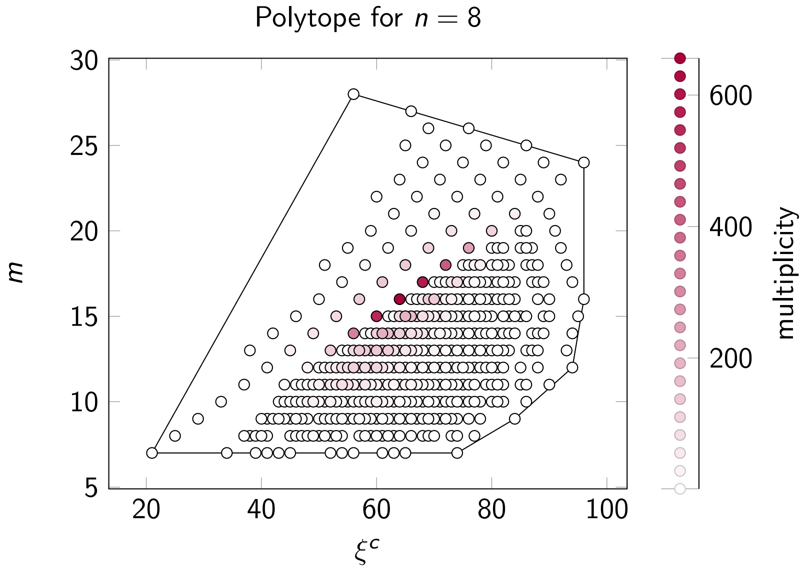
Exploring ξ^c with PHOEG: polytopes



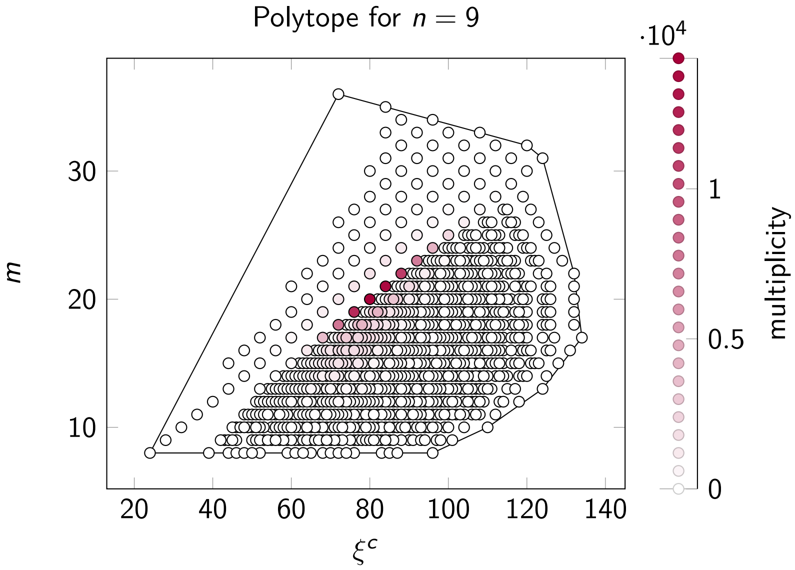
Exploring ξ^c with PHOEG: polytopes



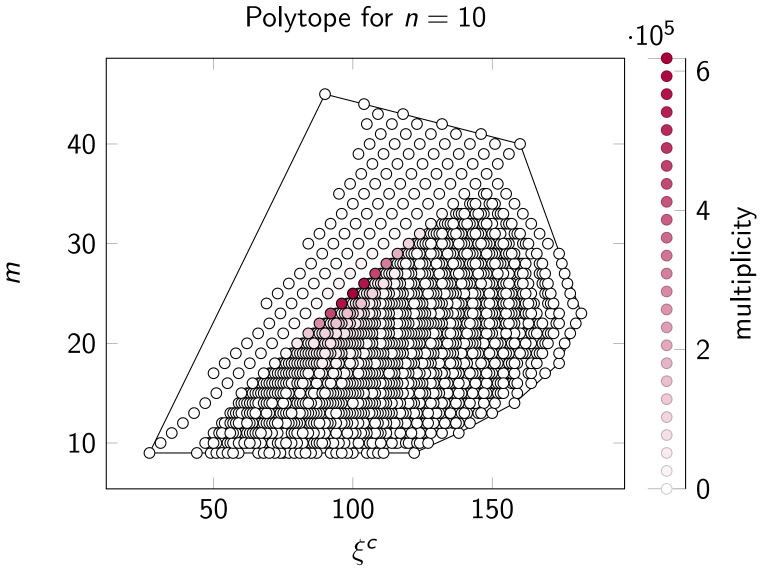
Exploring ξ^c with PHOEG: polytopes



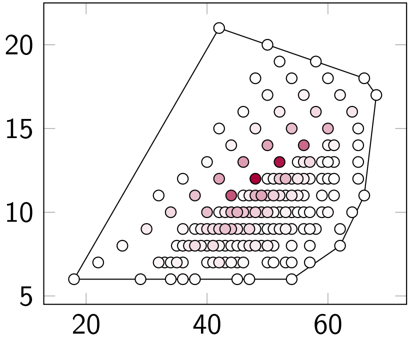
Exploring ξ^c with PHOEG: polytopes



Exploring ξ^c with PHOEG: polytopes



Observations and questions



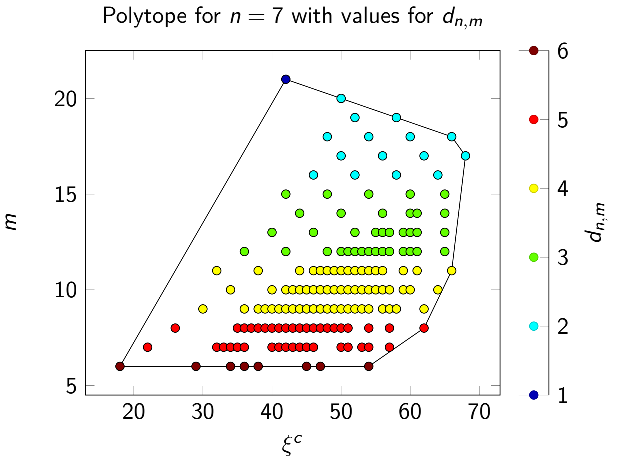
- How to explain the grid?
- Is the conjecture of Zhang, Liu and Zhou true when $d_{n,m} \geq 3$?
- Upper bound when $d_{n,m} < 3$?

Database query – Polytope with some 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 $d_{n,m}$



Recall that the conjecture is stated for $d_{n,m} \geq 3$. Is it true for $n = 7$?

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
```

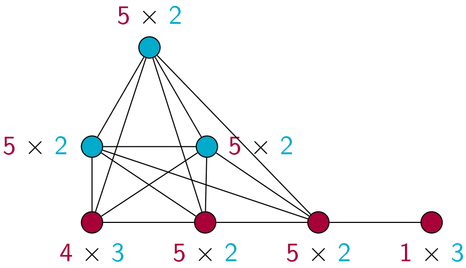
Database query – Extremal graphs

```
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;
```

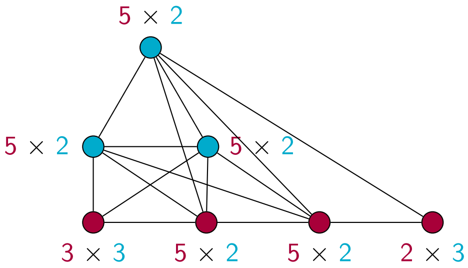
```
 sig   | n | m  | eci | d
-------+---+----+-----+----
F@IQO  | 7 |  6 |  54 | 6
F@‘J_  | 7 |  7 |  57 | 5
FgCXW  | 7 |  8 |  62 | 5
FWCYw  | 7 |  9 |  62 | 4
FgCxw  | 7 | 10 |  64 | 4
F‘Kyw  | 7 | 11 |  66 | 4
F‘Kzw  | 7 | 12 |  65 | 3
F‘Lzw  | 7 | 13 |  65 | 3
F‘\zw  | 7 | 14 |  65 | 3
FJ]|w  |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

⇒ counter-example to the conjecture !
Extremal graphs are not always unique

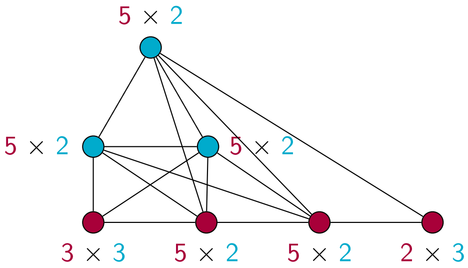
Counter-example ($n = 7$ and $m = 15$)



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

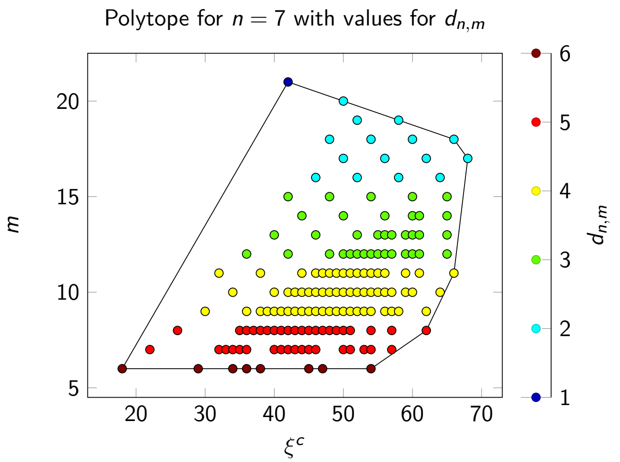


Counter-example ($n = 7$ and $m = 15$)



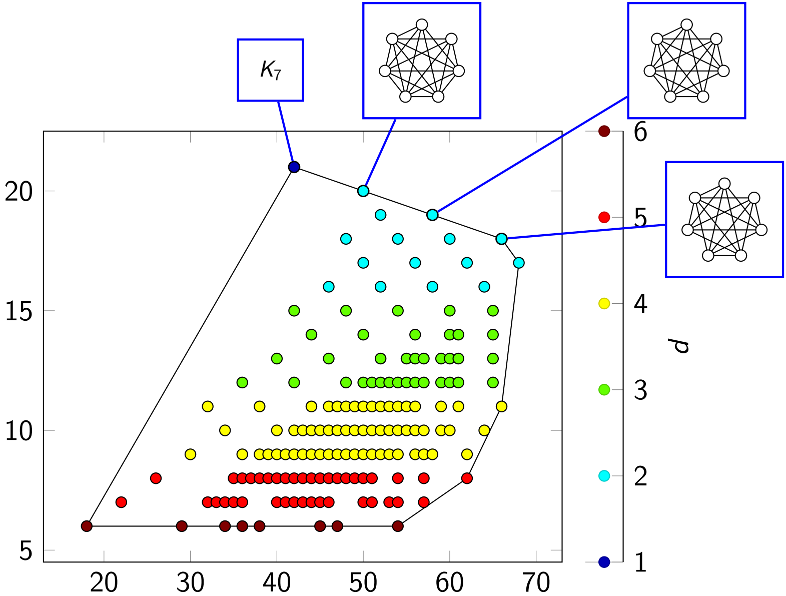
It is possible to construct counter-examples for any values of $n \geq 6$ (with $d_{n,m} = 3$).

Coloring points with values of $d_{n,m}$

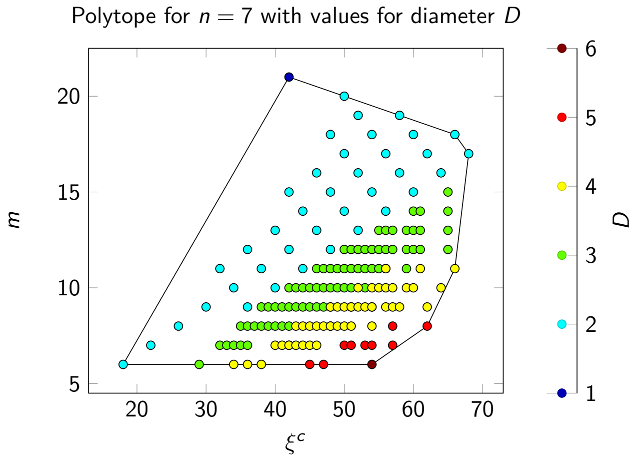


Upper bound when $d_{n,m} < 3$?

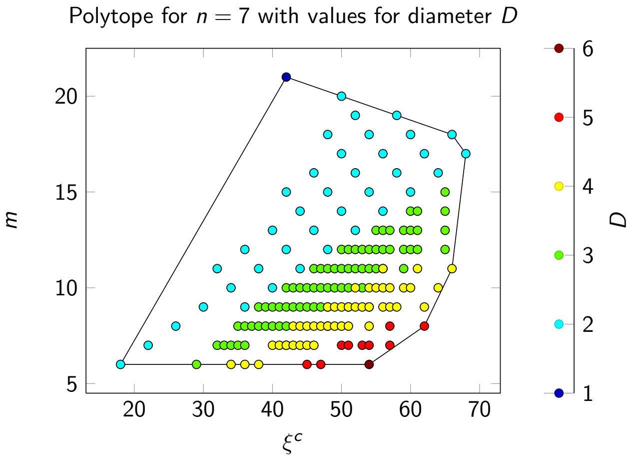
Upper facet of the polytope ($n = 7$)



Coloring points with values of the diameter



Coloring points with values of the diameter



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

A new tight upper bound when $d_{n,m} < 3$

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 $d_{n,m} < 3$).

Number of non-equivalent colorings

We note $P(G, k)$ the number of non-equivalent colorings of G that use exactly k colors.

$$P(P_3, 2) = 1$$

$$P(P_3, 3) = 1$$

Total number of non-equivalent colorings

Definition

The total number of non-equivalent colorings $\mathcal{P}(G)$ of a graph G is

$$\mathcal{P}(G) = \sum_{k=0}^{n} \mathrm{P}(G,k) = \sum_{k=\chi(G)}^{n} \mathrm{P}(G,k),$$

where $\chi(G)$ is the chromatic number of G.

Example: $\mathcal{P}(\mathrm{P}_3) = \mathrm{P}(\mathrm{P}_3, 2) + \mathrm{P}(\mathrm{P}_3, 3) = 1 + 1 = 2$

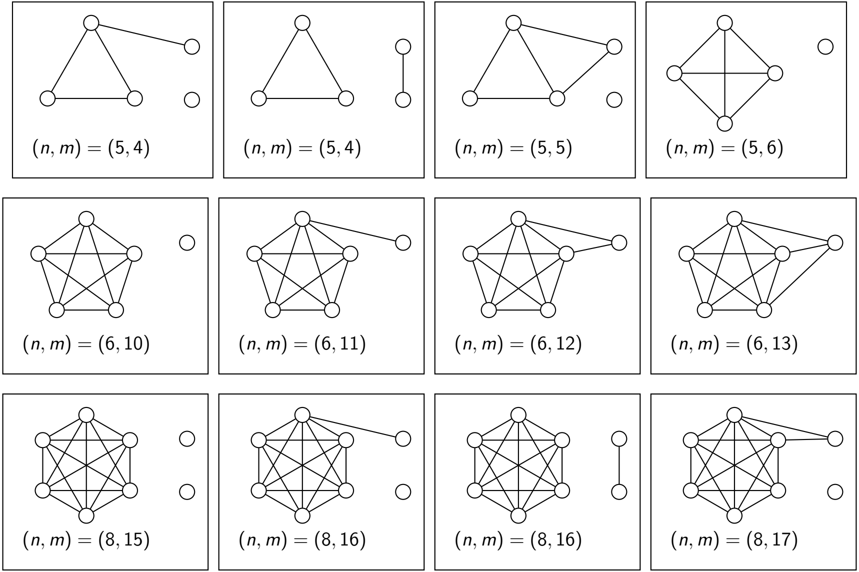
$\mathcal{P}(G)$ is the value of the σ-polynomial when $x = 1$ and is also known as the Bell number of a graph [Duncan & Peele, 2009].

The Min-NumCol-NumEdges Problem

Problem

What is minimum possible value of $\mathcal{P}$ for graphs of fixed order n and size m and what are the graphs attaining those bounds ?

Some extremal graphs



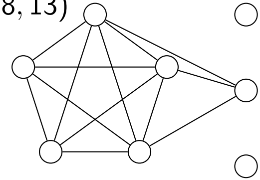
The extremal(?) graphs

Given n the order and m the size of graphs. Let t_k be the biggest triangular number such that $t_k \leq m$. We call $r_m = m - t_k$ the remainder.

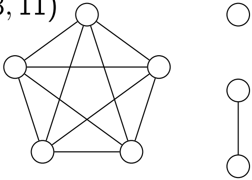
We define $G^*(n, m)$ as the unique graph formed from $K_{k+1} \bigcup \overline{K}_{n-k-1}$, where one (if any) vertex of $\overline{K}_{n-k-1}$ is connected to r_m vertices of the clique.

If $r_m = 1$, and $n - k - 1 \geq 2$, we define $G'(n, m)$ as $K_{k+1} \bigcup \overline{K}_{n-k-1}$, where two vertices of $K_{k+1} \bigcup \overline{K}_{n-k-1}$ are connected.

$G^*(8, 13)$

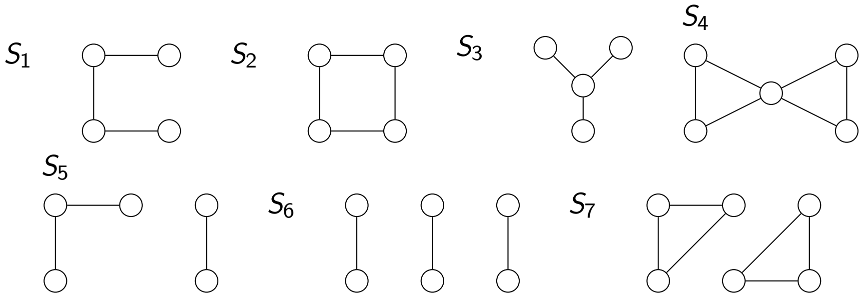


$G'(8, 11)$



Forbidden Graph Characterization

In this tool, we want a necessary *and* sufficient characterization of our graphs.



Concluding remarks

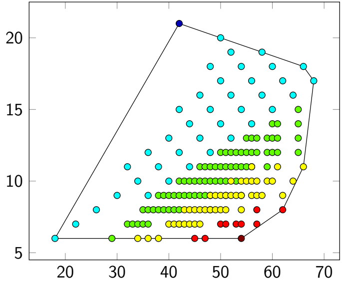
- Not only extremal graphs are useful to study extremal properties of an invariant
- Exact approach limited to small graphs ($n \leq 10$)
- However, dealing with small graphs has already shown to be very useful and led to numerous results (AutoGraphiX, GraPHedron)

Perspectives

- Invariants' DB allows a form of dynamic programming;
- Create a simple interface for queries, define a domain specific language;
- Allow easy visualization and manipulation of outputs (GUI, PDF, etc.);
- Go up in the order of graphs, relaxing the *exact* constraint.

Appendix

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.

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

Definition

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

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

In the following, we simply use d for $d_{n,m}$.

Definition

Let $E_{n,m}$ be the graph obtained from a clique K_{n-d-1} and a path $P_{d+1} = v_0 v_1 \ldots v_d$ by joining each vertex of the clique to both v_d and v_{d-1}, and by joining

$$m-n+1-\binom{n-d}{2}$$

vertices of the clique to v_{d-2}.

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

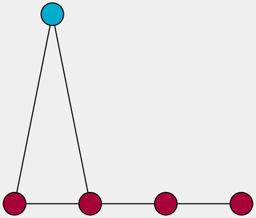
Example ($n = 5$)

m	**4**	5	6	7	8	9	10
d	**4**	3	3	2	2	2	1
$n - d - 1$	**0**	1	1	2	2	2	3
# edges to v_{d-2}	**0**	0	1	0	1	2	0

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

Example ($n = 5$)

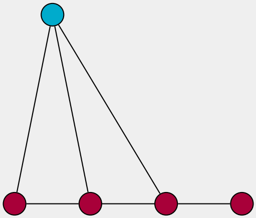
m	4	**5**	6	7	8	9	10
d	4	**3**	3	2	2	2	1
$n-d-1$	0	**1**	1	2	2	2	3
# edges to v_{d-2}	0	**0**	1	0	1	2	0



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

Example ($n = 5$)

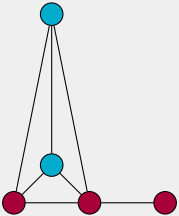
m	4	5	**6**	7	8	9	10
d	4	3	**3**	2	2	2	1
$n - d - 1$	0	1	**1**	2	2	2	3
# edges to v_{d-2}	0	0	**1**	0	1	2	0



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

Example ($n = 5$)

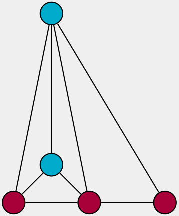
m	4	5	6	**7**	8	9	10
d	4	3	3	**2**	2	2	1
$n - d - 1$	0	1	1	**2**	2	2	3
# edges to v_{d-2}	0	0	1	**0**	1	2	0



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

Example ($n = 5$)

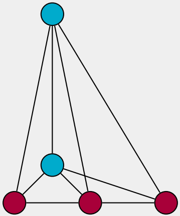
m	4	5	6	7	**8**	9	10
d	4	3	3	2	**2**	2	1
$n - d - 1$	0	1	1	2	**2**	2	3
# edges to v_{d-2}	0	0	1	0	**1**	2	0



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

Example ($n = 5$)

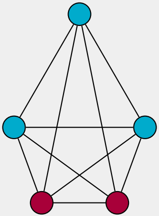
m	4	5	6	7	8	**9**	10
d	4	3	3	2	2	**2**	1
$n - d - 1$	0	1	1	2	2	**2**	3
# edges to v_{d-2}	0	0	1	0	1	**2**	0



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

Example ($n = 5$)

m	4	5	6	7	8	9	**10**
d	4	3	3	2	2	2	**1**
$n - d - 1$	0	1	1	2	2	2	**3**
# edges to v_{d-2}	0	0	1	0	1	2	**0**



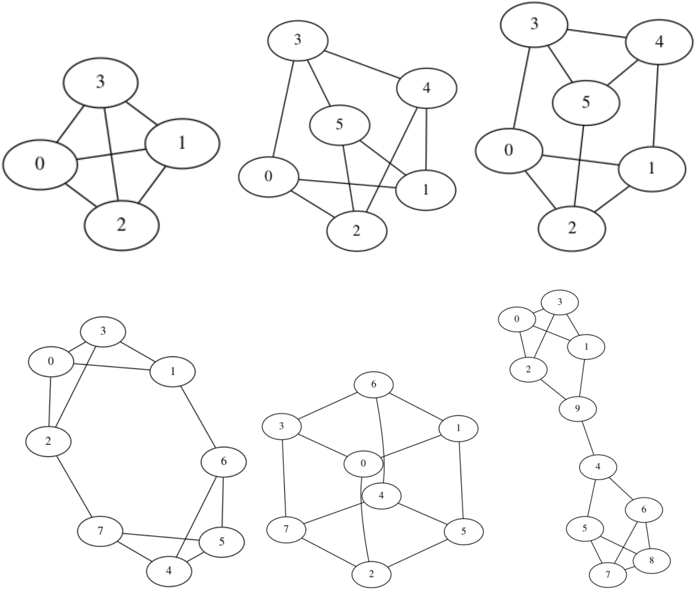
What about other classes of graphs ?

Let's try to maximize ξ^c on cubic (3-regular) graphs.

```
SELECT t.n, t.signature, t.eci
FROM (
  SELECT n.val AS n, eci.signature, eci.val as eci,
    DENSE_RANK() OVER (
      PARTITION BY n.val
      ORDER BY eci.val DESC
    ) AS pos
  FROM cubic
  JOIN num_vertices n USING(signature)
  JOIN eccentric_connectivity_index eci USING(signature)
  ) t
WHERE t.pos = 1
ORDER BY t.n;
```

Maximize ξ^c on cubic graphs

n	signature	eci
4	C~	12
6	Es\o	36
6	E{Sw	36
8	Gv?IXW	72
8	Gs@ipo	72
10	Iv?GOKFY?	126
12	Kt?GOKFOAOeA	177
14	Mt?GO?@@_KgKOWM??	270
16	Ot?G?CA?WB`oO?O?b_@?E	348
18	Qv??W[K?G??@?B?B?A??`G?p??o	474
20	Sv?GW?@?W??@?B????G?J??w?w?M?BO??	573
22	Uv?G?CK?oE@_?H?E??G?C??C??W?@??@C_?KO??o	726


3 0 1 2
3 4 5 0 1 2
3 4 5 0 1 2
3 0 1 2 6 7 5 4
6 3 1 0 4 7 5 2
3 0 1 2 9 4 6 5 8 7