House of Graphs: what are interesting graphs?

G. Brinkmann\textsuperscript{1}, K. Coolsaet\textsuperscript{1}, J. Goedgebeur\textsuperscript{1} and H. Mélot\textsuperscript{1,2}

\textsuperscript{1} Vakgroep Toegpaste Wiskunde en Informatica, UGent, Belgium
\textsuperscript{2} Service d’Informatique Théorique, UMons, Belgium

21 July 2010
Computers in Scientific Discovery 5, Sheffield, UK
Objectives

Main objectives of the House of Graphs project:

- What make a graph relevant or interesting?
- Amongst the large number of non isomorphic graphs, is there a few that can be considered as interesting?
- How to share the answers of the two previous questions with researchers?
**Notations**

**Definition**
A graph $G = (V,E)$:

- set $V$ of **nodes**;
- set $E$ of **edges**.

**Remark**
Graphs considered:
**simples and undirected**
**Notations**

**Definition**
A graph **invariant** is a numerical value, preserved by isomorphism.

**Example**
Numbers $n$ of nodes and $m$ of edges.

**Example**
$n = 4$ et $m = 5$. 
What make a graph relevant of interesting?

We propose two answers:

- appears useful in the literature or in (static) websites;
- is pointed out by a conjecture-making system.

Examples: complete graphs, cycles, paths, Petersen graph, Heawood graph (cf. Pisanski’s talk), etc.
Interesting graphs in the literature or on the web

Interesting graphs in the literature and on the web: counterexamples; tight graphs; classes of graphs, lists of graphs, etc.
Interesting graphs in the literature or on the web

Interesting graphs in the literature and on the web: counterexamples; tight graphs; classes of graphs, lists of graphs, etc.

Examples of books:

- Capobianco, Molluzzo, *Examples and Counterexamples in Graph Theory* (1978)
Interesting graphs in the literature or on the web

Interesting graphs in the literature and on the web: counterexamples; tight graphs; classes of graphs, lists of graphs, etc.

Examples of books:

- Capobianco, Molluzzo, *Examples and Counterexamples in Graph Theory* (1978)
Interesting graphs in the literature or on the web

Interesting graphs in the literature and on the web: counterexamples; tight graphs; classes of graphs, lists of graphs, etc.

### Graphs

This page contains some collections of graphs. See the combinatorial data page.

#### Simple graphs

<table>
<thead>
<tr>
<th>Vertices</th>
<th>All</th>
<th>Connected</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>all (2)</td>
<td>connected (1)</td>
</tr>
<tr>
<td>3</td>
<td>all (4)</td>
<td>connected (2)</td>
</tr>
<tr>
<td>4</td>
<td>all (11)</td>
<td>connected (6)</td>
</tr>
<tr>
<td>5</td>
<td>all (34)</td>
<td>connected (21)</td>
</tr>
<tr>
<td>6</td>
<td>all (156)</td>
<td>connected (112)</td>
</tr>
<tr>
<td>7</td>
<td>all (1064)</td>
<td>connected (853)</td>
</tr>
<tr>
<td>8</td>
<td>all (12346)</td>
<td>connected (11117)</td>
</tr>
<tr>
<td>9</td>
<td>all (274668)</td>
<td>connected (261080)</td>
</tr>
<tr>
<td>10</td>
<td>all (313088)</td>
<td>connected (308168)</td>
</tr>
</tbody>
</table>

The above graphs, and many varieties of them, can be efficiently generated using a table giving the number of graphs according to the number of edges and vertices.

#### Eulerian graphs

Here we give the small simple graphs with every degree even.

<table>
<thead>
<tr>
<th>Vertices</th>
<th>All</th>
<th>Connected</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>all (1)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>all (2)</td>
<td>connected (1)</td>
</tr>
<tr>
<td>4</td>
<td>all (3)</td>
<td>connected (1)</td>
</tr>
<tr>
<td>5</td>
<td>all (7)</td>
<td>connected (4)</td>
</tr>
<tr>
<td>6</td>
<td>all (16)</td>
<td>connected (8)</td>
</tr>
<tr>
<td>7</td>
<td>all (54)</td>
<td>connected (37)</td>
</tr>
<tr>
<td>8</td>
<td>all (243)</td>
<td>connected (184)</td>
</tr>
<tr>
<td>9</td>
<td>all (508)</td>
<td>connected (1792)</td>
</tr>
<tr>
<td>10</td>
<td>all (33120)</td>
<td>connected (31026)</td>
</tr>
<tr>
<td>11</td>
<td>all (1182004)</td>
<td>connected (1148626)</td>
</tr>
<tr>
<td>12</td>
<td>part 1, part 2, part 3, part 4</td>
<td>(each file about 81MB) (87223296)</td>
</tr>
</tbody>
</table>

### Examples of books:

- Capobianco, Molluzzo, *Examples and Counterexamples in Graph Theory* (1978)

### Examples of websites (static lists of graphs):

- Brendan McKay
- Markus Meringer
- Gordon Royle
Interesting graphs in the literature or on the web

Interesting graphs in the literature and on the web: counterexamples; tight graphs; classes of graphs, lists of graphs, etc.

Examples of books:

- Brandstädt, Le and Spinrad, Graph classes: a survey (1999)
- Capobianco, Molluzzo, Examples and Counterexamples in Graph Theory (1978)

Examples of websites (static lists of graphs):

- Brendan McKay
- Markus Meringer
- Gordon Royle
Conjecture-making systems

For particular problems (conjectures, set of invariants, inequality of invariants, etc.), graphs are pointed out by conjecture-making systems.

Examples:
- **AutoGraphiX**: extremal graphs;
- **GrInvIn**: counterexamples;
- **Graffiti**: counterexamples;
- **GraPHedron**: vertex-graphs (= “conglomerates”, see later);
- new version of **newGRAPh**? see Friday…
- etc.
Is there a few graphs that are interesting?

Amongst the large number of non isomorphic graphs, is there a few that can be considered as interesting?

Our hypothesis: very few graphs can be considered as interesting.
A first definition of interesting graphs

Starting point to obtain (automatically) a first set of interesting graphs: use of GraPHedron.

GraPHedron\(^1\)

- Computer assisted and automated conjectures
- Use a polyhedral approach
- Conjectures (inequalities among graph’s invariants) : best possible under some conditions

\(^1\)HM, Disc. Appl. Math. 156 (2008), 1875-1891
GraPHedron’s type of problems

Definition

A problem is defined by $I$, $C$ and $n$, where

- $I = (f, g)$ is a pair of graph’s invariants $f$ and $g$ (excluding the number of nodes $n$);
- $C$ is a particular class of graphs;
- $n$ is a fixed number of nodes.

Problem

What are all the best linear inequalities among $f$ and $g$, valid for all graphs of order $n$ in $C$?
GraPHedron’s type of problems

**Input**: a problem defined by $I = (f, g)$, $C$ and $n$.

**Output**: a polyhedral description (polytope $\mathcal{P}$) of the problem

$$\mathcal{P} = \text{conv}\{(x, y) \mid \exists G = (V, E) \in C, |V| = n, f(G) = x, g(G) = y\}.$$

**Remarks**

- In this framework, we limit $I$ to 2 invariants (not the case in GraPHedron);
- In this talk: $C$ will be either general or connected graphs.
Polyhedral approach

Example: diameter $D$ and number of edges $m$ of connected graphs

$I = (D, m)$ and $C = \text{connected}$
Polyhedral approach
Example: diameter $D$ and number of edges $m$ of connected graphs

1. **generate** graphs $\in C_n$

Example ($n = 4$)

$I = (D, m)$ and $C = \text{connected}$

Graphs of $C_4$:

```
\begin{tikzpicture}
  \foreach \i in {1,...,4} {
    \node (v\i) at (90 - \i*90/4:1) {};
    \foreach \j in {\i,...,4} {
      \draw (v\i) -- (v\j) if \i < \j then \else if \i > \j then \fi \fi;
    }
  }
\end{tikzpicture}
```
Polyhedral approach

Example: diameter $D$ and number of edges $m$ of connected graphs

1. generate graphs $\in C_n$
2. compute **invariants** of $I$

Example $(n = 4)$

$I = (D, m)$ and $C =$ connected

Graphs of $C_4$:

Coordinates $(D, m)$

- $(2,3)$
- $(3,3)$
- $(2,4)$
- $(2,5)$
- $(1,6)$

$D + m \leq 7$, $2D + m \leq 9$, $m \geq 3$, $3D + m \geq 9$. 
Polyhedral approach
Example: diameter $D$ and number of edges $m$ of connected graphs

1. generate graphs $\in C_n$
2. compute invariants of $I$
3. consider graphs as **points** in the space

$I = (D, m)$ and $C = \text{connected}$
Polyhedral approach
Example: diameter $D$ and number of edges $m$ of connected graphs

1. generate graphs $\in C_n$
2. compute invariants of $I$
3. consider graphs as points in the space
4. compute the polytope $P$ (convex hull)

\[ I = (D, m) \text{ and } C = \text{connected} \]
Polyhedral approach

Example: diameter $D$ and number of edges $m$ of connected graphs

1. generate graphs $\in C_n$
2. compute invariants of $I$
3. consider graphs as points in the space
4. compute the polytope $P$ (convex hull)
5. **Facets** of $P_n$: linear inequ. among $I$

Example ($n = 4$)

$I = (D, m)$ and $C = \text{connected}$

```
1 3 2
D
m
1
2
3
4
5
6
```

$D + m \leq 7$, $2D + m \leq 9$, $m \geq 3$, $3D + m \geq 9$.
Polyhedral approach
Example: diameter $D$ and number of edges $m$ of connected graphs

1. generate graphs $\in C_n$
2. compute invariants of $I$
3. consider graphs as points in the space
4. compute the polytope $P$ (convex hull)
5. Facets of $P_n$: linear inequ. among $I$

**Example ($n = 4$)**

$I = (D, m)$ and $C = \text{connected}$

\[
D + m \leq 7,
2D + m \leq 9,
m \geq 3,
3D + m \geq 9.
\]
Polyhedral approach

**Facet Defining Inequalities** (FDI) of $\mathcal{P}$ are “all the best” linear inequalities among $I$:
- cannot be deduced from other valid inequalities
- constitute a minimal system describing the polytope

$\Rightarrow$ useful for conjecture-making in the GPH framework

**Graphs that are pointed out**: graphs that correspond to the vertices of $\mathcal{P}$.

$\Rightarrow$ useful in the current framework
First definition of interesting graphs

Definition

A **vertex-graph** is a graph whose corresponding point in the space of invariants is a vertex of \( \mathcal{P} \).

Example \((I = (D, m), \mathcal{C} = \text{connected}, n = 4)\)

Vertex-graphs:

- star \( S_4 \)
- path \( P_4 \)
- complete graph \( K_4 \)
- graph \( K_4 \setminus e \)

Vertex-graphs are **interesting** as they are extremal for the problem but...
A vertex-graph can be considered as interesting for a problem but there can be a lot of graphs sharing the same pair of coordinates.

**Definition**

A conglomerate is a set of vertex-graphs that have the same pair of coordinates, for a given problem.
First definition of interesting graphs

Conglomerates

Example

Let $\delta$ be the minimum degree.

Problem definition: $I = (m, \delta); C =$ connected; $n = 9$.

If $T$ is a tree, then its minimum degree $\delta$ is 1 and its number of edges $m$ is $n - 1$.

$\Rightarrow$ all 47 trees with 9 nodes form a conglomerate with coordinates $(8, 1)$.
First definition of interesting graphs

Conglomerates

Example

Amongst the 12005168 non isomorphic graphs with 10 nodes, 11286671 graphs (94.01\%) have a matching number equals to 5 and a number of dominant nodes\(^a\) equals to 0.

\(^a\)node with a degree equals to \(n - 1\)
First definition of interesting graphs

Conglomerates

Example

Amongst the 12005168 non isomorphic graphs with 10 nodes, 11286671 graphs (94.01%) have a matching number equals to 5 and a number of dominant nodes\(^a\) equals to 0.

\(^a\)node with a degree equals to \(n - 1\)

- The fact that a graph is a vertex-graph for a problem is not sufficient to define it as interesting;
- However, graphs in a conglomerate can be considered as similar for a given problem (they share some properties);
- For example, using stars and paths is often sufficient to work on a conjecture about trees.
First definition of interesting graphs

Minimum set of covering graphs

We refine the definition of interesting graphs.

**Definition**

Let $C$ be a set of conglomerates. The set of interesting graphs induced by $C$ is the minimum set of graphs that cover all conglomerates of $C$. 

▶ finding the minimum set of graphs is equivalent to the MINIMUM SET COVER problem (NP-hard) 

▶ no hope to have an (efficient) exact algorithm (unless $P=NP$) 

▶ a greedy heuristic is known as the best-possible (polynomial) approximation algorithm for this problem
First definition of interesting graphs

Minimum set of covering graphs

We refine the definition of interesting graphs.

Definition

Let \( C \) be a set of conglomerates. The set of interesting graphs induced by \( C \) is the minimum set of graphs that cover all conglomerates of \( C \).

- finding the minimum set of graphs is equivalent to the MINIMUM SET COVER problem (NP-hard)
- no hope to have an (efficient) exact algorithm (unless \( P = NP \))
- a greedy heuristic is known as the best-possible (polynomial) approximation algorithm for this problem
Computational results

Two classes of graphs:

- connected: 23 invariants (\(\approx 253\) problems)

- average degree;
- average distance;
- chromatic number;
- clique number;
- cycle rank;
- diameter;
- edge connectivity;
- Fibonacci index;
- forest number;
- irregularity;
- maximum degree;
- matching number;
- minimum degree;
- minimum vertex cover;
- number of pendant nodes;
- number of nodes with degree \(n - 1\);
- number of edges;
- proximity;
- radius;
- remoteness;
- stability number;
- variance of degrees;
- variance of distances;
Computational results

Two classes of graphs:

- connected: 23 invariants (≈ 253 problems)
- general: 17 invariants (≈ 136 problems)

- average degree;
- average distance;
- chromatic number;
- clique number;
- cycle rank;
- diameter;
- edge connectivity;
- Fibonacci index;
- forest number;
- irregularity;
- maximum degree;
- matching number;
- minimum degree;
- minimum vertex cover;
- number of pendant nodes;
- number of nodes with degree $n - 1$;
- number of edges;
- proximity;
- radius;
- remoteness;
- stability number;
- variance of degrees;
- variance of distances;
Computational results

Total: 389 problems (for each value of $n = 4, 5, \ldots, 10$).

<table>
<thead>
<tr>
<th>$n$</th>
<th># gr.</th>
<th># pol. vert.</th>
<th># cong.</th>
<th># int. gr.</th>
<th>pc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>11</td>
<td>1402</td>
<td>63</td>
<td>11</td>
<td>100.00%</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>1602</td>
<td>126</td>
<td>25</td>
<td>73.53%</td>
</tr>
<tr>
<td>6</td>
<td>156</td>
<td>1751</td>
<td>176</td>
<td>46</td>
<td>29.49%</td>
</tr>
<tr>
<td>7</td>
<td>1044</td>
<td>1932</td>
<td>236</td>
<td>73</td>
<td>6.99%</td>
</tr>
<tr>
<td>8</td>
<td>12346</td>
<td>2039</td>
<td>242</td>
<td>89</td>
<td>0.72%</td>
</tr>
<tr>
<td>9</td>
<td>274668</td>
<td>2253</td>
<td>320</td>
<td>127</td>
<td>0.05%</td>
</tr>
<tr>
<td>10</td>
<td>12005168</td>
<td>2338</td>
<td>323</td>
<td>168</td>
<td>0.001%</td>
</tr>
</tbody>
</table>

- # gr.: number of non isomorphic graphs
- # pol. vert.: total number of vertices for all polytopes
- # cong.: number of distinct conglomerates
- # int. gr.: number of interesting graphs (approx. by greedy heuristic)
- pc.: percent of interesting graphs
Computational results

Total: 389 problems (for each value of $n = 4, 5, \ldots, 10$).

<table>
<thead>
<tr>
<th>$n$</th>
<th># gr.</th>
<th># cong.</th>
<th># $\geq 2$</th>
<th># $\geq 5%$</th>
<th># $\geq 10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>11</td>
<td>63</td>
<td>45</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>126</td>
<td>92</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>156</td>
<td>176</td>
<td>109</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>1044</td>
<td>236</td>
<td>151</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>12346</td>
<td>242</td>
<td>153</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>274668</td>
<td>320</td>
<td>198</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>12005168</td>
<td>323</td>
<td>209</td>
<td>19</td>
<td>10</td>
</tr>
</tbody>
</table>

▶ # gr.: nbr of non isomorphic graphs
▶ # cong.: nbr of distinct conglomerates
▶ # $\geq 2$: nbr of dist. cong. that appears in at least 2 problems
▶ # $\geq 5\%$: nbr of dist. cong. that appears in at least 5% of the problems
▶ # $\geq 10\%$: nbr of dist. cong. that appears in at least 10% of the problems
Computational results

When $n = 6$ (similar for other values), the most popular graphs are...

- always in a conglomerate of size 1;
- all other conglomerates appears in less than 15% of the problems.
Current features of the prototype (see demonstration):

- 4 types of queries about interesting graphs (including interest relations and filters)
- refinable search
- static lists of particular graphs
- information about graphs and conglomerates
- results can be downloaded (graph6 or multicode)
The House of Graphs

http://hog.grinvin.org

⇒ demonstration
Perspectives

- GraPHedron’s type of interest:
  - compute more data with GraPHedron (more invariants, more classes)
  - recognition of conglomerates (rules, names, etc.)

- Add other definitions of interesting graphs:
  - use other conjecture-making systems
  - literature (difficult: not automatically)
  - user defined interesting graphs (should define policies and roles)

- add more information about graphs in the database (names, types of interest, etc.)

- what do you expect / find useful for such a tool?
House of Graphs: what are interesting graphs?

G. Brinkmann\(^1\), K. Coolsaet\(^1\), J. Goedgebeur\(^1\) and H. Mélot\(^{1,2}\)

\(^1\) Vakgroep Toegpast Wiskunde en Informatica, UGent, Belgium
\(^2\) Service d’Informatique Théorique, UMons, Belgium

21 July 2010
Computers in Scientific Discovery 5, Sheffield, UK