House of Graphs: what are interesting graphs?

G. Brinkmann[1], K. Coolsaet[1],
J. Goedgebeur[1] and H. Mélot[1,2]

[1] Vakgroep Toegpaste Wiskunde en Informatica, UGent, Belgium
[2] Service d'Informatique Théorique, UMons, Belgium

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

UNIVERSITEIT GENT

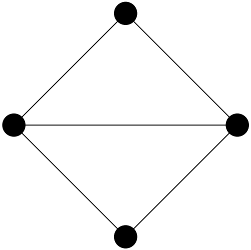
UMONS Université de Mons

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

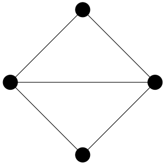
A graph **invariant** is a numerical value, preserved by isomorphism.

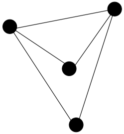
Example

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.

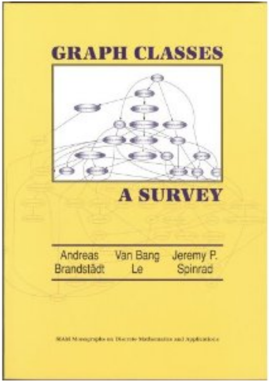
Examples: complete graphs, cycles, paths, Petersen graph, Heawood graph (cf. Pisanski's talk), etc.

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.

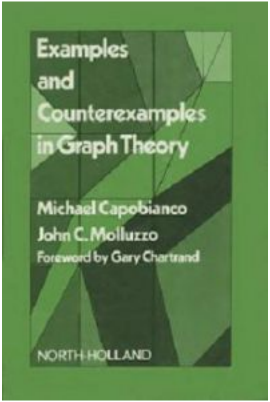


Examples of books:

- Brandstädt, Le and Spinrad, Graph classes: a survey (1999)
- 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:

- Brandstädt, Le and Spinrad, Graph classes: a survey (1999)
- 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

2 vertices: all (2) connected (1)
3 vertices: all (4) connected (2)
4 vertices: all (11) connected (6)
5 vertices: all (34) connected (21)
6 vertices: all (156) connected (112)
7 vertices: all (1044) connected (853)
8 vertices: all (12346) connected (11117)
9 vertices: all (274668) connected (261080)
10 vertices: all (31MB gzipped) (12005168) connected (30MB gzipped) (1171

The above graphs, and many varieties of them, can be efficiently generated usin

A table giving the number of graphs according to the number of edges and vertic

Eulerian graphs

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

2 vertices: all (1)
3 vertices: all (2) connected (1)
4 vertices: all (3) connected (1)
5 vertices: all (7) connected (4)
6 vertices: all (16) connected (8)
7 vertices: all (54) connected (37)
8 vertices: all (243) connected (184)
9 vertices: all (2038) connected (1782)
10 vertices: all (33120) connected (31026)
11 vertices: all (1182004) connected (1148626)
12 vertices: part 1; part 2; part 3; part 4; (each file about 81MB) (87723296)

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

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.

Connected regular graphs

The following table contains numbers of connected regular graphs with given number of vertices a degree. For the empty fields the number is not yet known (to me). The latest numbers (for n=19,k= n=16,k=6; n=16,k=7) have been contributed by Jason Kimberley (University of Newcastle, Austra 2009), who ran GENREG on up to 250 cores.

Vertices	Degree 3	Degree 4	Degree 5	Degree 6	Degree 7
4	1	0	0	0	0
5	0	1	0	0	0
6	2	1	1	0	0
7	0	2	0	1	0
8	5	6	3	1	1
9	0	16	0	4	0
10	19	59	60	21	5
11	0	265	0	266	0
12	85	1544	7848	7849	1547
13	0	10778	0	367860	0
14	509	88168	3459383	21609300	21609301
15	0	805491	0	1470293675	0
16	4060	8037418	2585136675	113314233808	733351105934
17	0	86221634	0		0
18	41301	985870522			
19	0	11946487647			
20	510489				
22	7319447				
24	117940535				
26	2094480864				

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, $\mathcal{C}$ and n, where

- $I = (f, g)$ is a pair of graph's invariants f and g (excluding the number of nodes n);
- $\mathcal{C}$ 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 $\mathcal{C}$?

GraPHedron's type of problems

Input: a problem defined by $I = (f, g)$, $\mathcal{C}$ and n.

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

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

Remarks

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

Polyhedral approach

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

Example ($n = 4$)

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

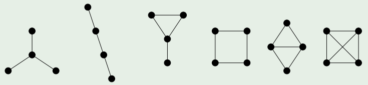
Polyhedral approach

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

1. **generate** graphs $\in \mathcal{C}_n$

Example ($n = 4$)

$\mathcal{I} = (D, m)$ and $\mathcal{C} =$ connected
Graphs of $\mathcal{C}_4$:



Polyhedral approach

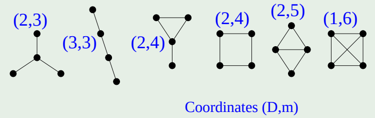
Example: diameter *D* and number of edges *m* of connected graphs

1. generate graphs $\in \mathcal{C}_n$
2. compute **invariants** of I

Example ($n = 4$)

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

Graphs of $\mathcal{C}_4$:



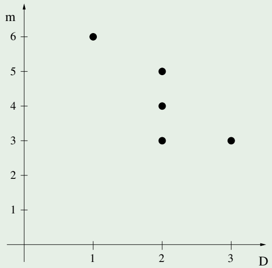
Polyhedral approach

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

1. generate graphs $\in \mathcal{C}_n$
2. compute invariants of I
3. consider graphs as **points in the space**

Example ($n = 4$)

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



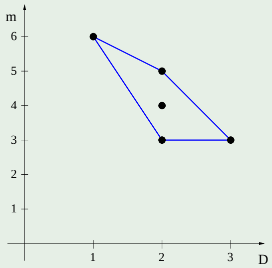
Polyhedral approach

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

1. generate graphs $\in \mathcal{C}_n$
2. compute invariants of I
3. consider graphs as points in the space
4. compute the **polytope** $\mathcal{P}$ (convex hull)

Example ($n = 4$)

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



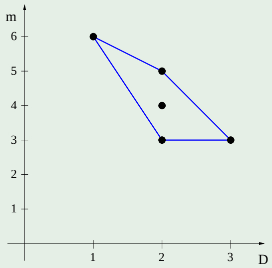
Polyhedral approach

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

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

Example ($n = 4$)

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



Polyhedral approach

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

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

Example ($n = 4$)

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

$$\begin{aligned} D + m &\leq 7, \\ 2D + m &\leq 9, \\ m &\geq 3, \\ 3D + m &\geq 9. \end{aligned}$$

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

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

Graphs that are pointed out: graphs that correspond to the vertices of $\mathcal{P}$.
$\Longrightarrow$ useful in the current framework

First definition of interesting graphs

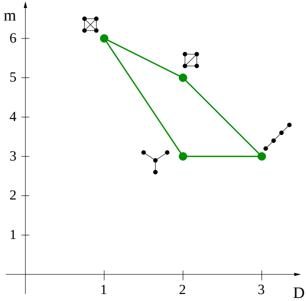
Definition

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

Example ($I = (D, m)$, $\mathcal{C} =$ 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. . .

First definition of interesting graphs

Conglomerates

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.

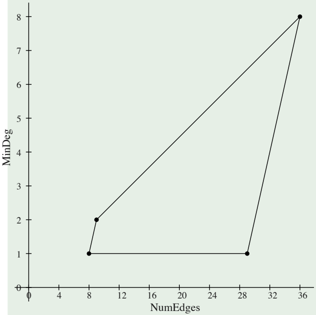
First definition of interesting graphs

Conglomerates

Example

Let δ be the minimum degree.

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



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

$\Longrightarrow$ 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.

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 ($= 253$ 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 $n = 4, 5, \ldots, 10$).

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

- # 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

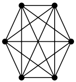
Total: 389 problems (for each value of $n = 4, 5, \ldots, 10$).

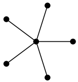
n	# gr.	# cong.	# ≥ 2	# $\geq 5\%$	# $\geq 10\%$
4	11	63	45	12	8
5	34	126	92	13	9
6	156	176	109	16	10
7	1044	236	151	14	9
8	12346	242	153	17	10
9	274668	320	198	19	10
10	12005168	323	209	19	10

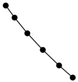
- # gr.: nbr of non isomorphic graphs
- # cong.: nbr of distinct conglomerates
- # ≥ 2: nbr of dist. cong. that appears in at least 2 problems
- # $\geq 5\%$: nbr of dist. cong. that appears in at least 5% of the problems
- # $\geq 10\%$: nbr of dist. cong. that appears in at least 10% of the problems

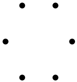
Computational results

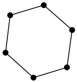
When $n = 6$ (similar for other values), the most popular graphs are...

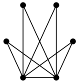

(94.1%)

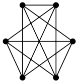

(57.8%)


(34.7%)


(28.5%)


(19.5%)


(17.2%)


(15.1%)

- always in a conglomerate of size 1;
- all other conglomerates appears in less than 15% of the problems.

The House of Graphs

http://hog.grinvin.org

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

the House of Graphs

http://hog.grinvin.org

$\Longrightarrow$ 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 Toegpaste Wiskunde en Informatica, UGent, Belgium
[2] Service d'Informatique Théorique, UMons, Belgium

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

UNIVERSITEIT GENT

UMONS
Université de Mons