A computer program for mathematical discovery

Ermelinda DeLaViña
University of Houston-Downtown
March 29, 2008

Some main practices of

a mathematician:

- Concept formation
- Conjecture making
- Constructing counterexamples
- Theorem proving

What makes a problem good?

- provokes thought,
- causes surprise,
- stimulates interest or
- inspires further research

In my abstract

- "One measure of a good conjecture or problem in mathematics is that it lead to new tools and techniques for making advances in the area under investigation. …."
- I have seen this attributed to problems of well known mathematicians.

Ramsey Theory

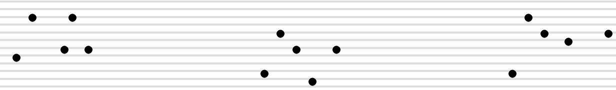
- Idea: Given a (mathematical) structure find the threshold on the order of the structure such that structures of that order admit a particular substructure.

Discovery & rediscovery of Ramsey Theory

- Schur (1916) & van der Waerden (1927) – Number theory
- Ramsey (1930) - Logic
- Dilworth (1950) – partially ordered sets
- Erdös & Szekeres (1935) – geometry & graph theory
- Erdös coined the term “Ramsey Theory”

How Erdös & Szekeres rediscovered Ramsey theory

- Ester Klein (later Szekeres) asked and solved the following simple problem:
- Given any five points on the plane no three of which are on a common line, is it true that four of them always form a convex quadrilateral?



- P. Erdös & G. Szekeres generalize this problem.
- Erdös popularized the graph theoretical version.

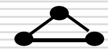
Graph theoretical version

Let *R(r,b)* be the smallest *n* such that any graph on *n* or more vertices has either:

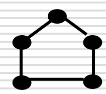
- a subset of *r* vertices every pair of which is adjacent (K_r subgraph), or
- a subset of *b* vertices every pair of which is not adjacent.

An example *R(3,3)*

K_3



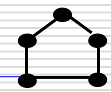
- *R(3,3)* is the smallest *n* such that any graph on *n* or more vertices has either 3 vertices that are pair-wise adjacent or 3 vertices that are pair-wise not adjacent.
- It is known and easily proved that R(3,3) = 6. This graph is called a critical graph.



new tools and techniques ?

- Erdös proved that $R(k,k) \geq 2^{k/2}$, introducing the idea of a random graph, and the probabilistic proof, which subsequently caught on and is applied in many areas.
- Algorithms for critical graph
- Variations- instead of subsets whose vertices are pair-wise adjacent or not, ask for guarantee of other substructures.

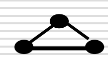
Bounds on the independence number



- Let G be a graph and let $\alpha(G)$ denote the **independence number of *G*** which is the order of a largest subset of the vertices of G such that no two are adjacent.
- Bounds on the independence number are of computational interest since computing it is NP-hard (no known polynomial-time algorithm)

Bounds on the independence number

K_3



- Now suppose that we consider only graphs that have no K_3 subgraph and that we happen to know a lower bound on $\alpha(G) \geq \frac{1}{3} n(G)$, where $n(G)$ is the number of vertices of G.
- Then K_3-free graphs with $\alpha(G) < b$ will have at most $3b$ vertices, and thus $R(3,b) \geq 3b$.

Graffiti in the 1980s

- A conjecture-making program created by Siemion Fajtlowicz in the mid 1980s.
- A recurrent theme in his series of papers *On Conjectures of Graffiti I*, II, III, IV. (1987-1990)* is the question,

  What makes a conjecture interesting?

1980s Graffiti question

- **Question**: Given a collection of graph invariants $\{inv_1, inv_2, \ldots, inv_k\}$, which relations of the form

  $inv_i \leq inv_j$, $\quad$ $inv_i \leq inv_j + 1$,

  $inv_i + inv_k \leq inv_j$, $\quad$ $inv_i \leq inv_j / inv_k$

… etc., are interesting?

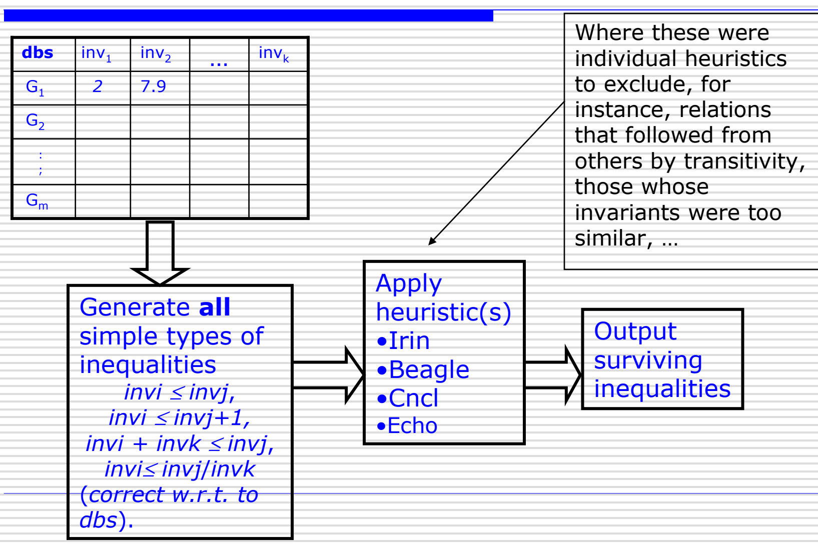
Graffiti's early conjectures

- Let *G* be a connected graph. Then
  - $\alpha(G) \geq$ radius of *G*
  - $\alpha(G) \geq$ average distance of *G*
  - $\alpha(G) \geq$ residue of *G*
- Each of the invariants on the right hand side are easily computable.
- They were novel at the time. The first two were subsequently proven multiple times.

reference

`http://www.math.uh.edu/~siemion`

1980s *Graffiti* strategy



Despite its success: Issues for & Drawbacks of 1980's Graffiti

- 1980's computing slow
- Computation of independence number and other NP-hard problems
- Limited expression types
- Limited number of invariants to be used since was generating all combinations.
- Hard to decide which heuristics or combinations of them worked best

reference

`http://www.math.uh.edu/~siemion`

Success?

- Conjectures engaged many researchers
- Resulted numerous research papers
- For bibliographical information on the papers see the URL below

reference

`http://cms.uhd.edu/faculty/delavinae/research/wowref.htm`

1990s *Graffiti* Notable changes

1. **algebraic form** of conjectures,
   - f_0 = I ={invariants},
     f_1={1/2, sqrt, ln, +1, -1, -1*, 2*, e^, 2^, reciprocal, …},
     f_2={+,×,^,…}

   T(1) = strings from <I, f_0, f_1, f_2>, which will be called *terms*.
2. Expressions generated and evaluated one at a time.
3. **Dalmatian Heuristic**
   - One heuristic to replace all but one of the previously used heuristics

Reference

S. Fajtlowicz, On Conjectures of Graffiti V (1995).

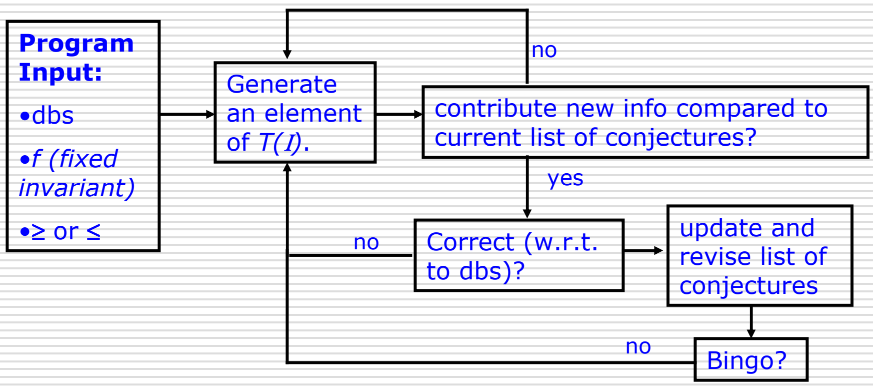
Dalmatian Heuristic

1. **Question**: Given a collection of graph invariants I, and a fixed term, *f* in *T(1)*, **which relations** $f(G) \geq c_i(G)$ where $c_i(G)$ is a term, **contribute new information** (compared to working list) about *f(G)*?
2. Moreover, he proposed to test for contribution of new information *before* testing for correctness (w.r.t to dbs). This at the time proved to be efficient.

Reference

S. Fajtlowicz, On Conjectures of Graffiti V (1995).

Dalmatian Heuristic



T(I) = strings from <I, *f*0, *f*1, *f*2>,

Dalmatian test for

- Given a list of conjectured bounds for a fixed term *f*,

  $f(G) \geq c_1(G)$,
  $f(G) \geq c_2(G)$, …
  $f(G) \geq c_k(G)$

  is it the case that there is at least one graph for which the new candidate $c_i(G)$ is closer to $f(G)$?

Bingo: Halting condition

- Given a list of conjectured bounds for a fixed term *f*,

  $f(G) \geq c_1(G)$,
  $f(G) \geq c_2(G)$, …
  $f(G) \geq c_k(G)$

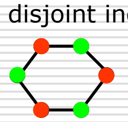
  is it the case that for each G in the dbs there exists a c_i such that $f(G) = c_i(G)$?

First Dalmatian Conjecture

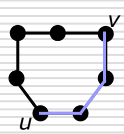
- 1992 Graffiti #747 for connected graphs $b/2 \geq$ *average distance.*

  where *b* is the number of vertices of a largest induced bipartite subgraph.

  A graph is ***bipartite*** if its vertices can be partitioned into 2 disjoint independent sets.



b/2 = 3
avg dist = 1.8



b/2 = 3
avg dist = 2

distance from u to v is the length of a shortest path from *u* to *v*.
dist(u,v) *= 3*

Average distance of a graph is the average of all distinct pairs of vertices.

reference

Written on the Wall: http://www.math.uh.edu/~siemion

747 recently proven

Theorem (DeLaVina & Waller)

$b/2 \geq$ *average distance - 0.5*

Theorem (Hansen et. al)

$f/2 \geq$ *average distance*

Note: $b \geq f$.

Of interest, in part,

because

Graffiti #747:

$b/2 \geq$ *average distance*

generalizes

Graffiti #2 (F. Chung 1988):

$\alpha \geq$ *average distance*

since $\alpha \geq b/2$.

My experience

In the early 1990's, as Fajtlowicz's research assistant I participated in the development Dalmatians heuristic of Graffiti and some ancillary tools.

Graffiti.pc

- In 2000-2001 a short-term goal was to have a Windows, Graffiti-like program that I could utilize with undergraduate students.
  - Numerous independent undergraduate research projects
  - In 2002 & 2004, Gunnar Brinkmann and Claudia Justus used Graffiti.pc for a graduate math education graph theory course and for a workshop for high school teachers of advanced students in Germany.
  - In 2007, Thomas Phaff at Ithica College used it for a proof based course.

Graffiti.pc

- By 2003, I had developed the program enough to use it for my own research.

Graffiti.pc

- Similar design as Graffiti.
  - A system of 2 programs
    - One builds the databases
    - The other generates conjectures
  - Initially only one conjecturing heuristic. A variation of Fajtlowicz's *Dalmatian*
  - In 2005, I added another heuristic dubbed *Sophie*, whose idea was inspired by discussion with Bill Waller.

Evaluating Conjectures

If G is connected, then

G.pc #13. $b \geq diam + \mu - 1$ — new & true

G.pc #14. $b \geq diam + freq(1) - 1$ — new & true

G.pc #15. $b \geq 2rad$ — previously known

G.pc #16. $b \geq 2(rad-1) + \mu$ — new & true

⋮

G.pc #21. $b \geq$ 2*avg dist. from boundary — new & still open

⋮

G.pc #27. $b \geq (\min \{|N(e)|: e \text{ an edge}\})^{1-t}$ — new & easily true

reference

http://cms.uhd.edu/faculty/delavinae/research/wowii

Necessary

- Dalmatian successfully produces *necessary conditions* for some specified property involving a fixed expression.
  - If *G* has property *P*, then *fixed(G)* ≥ *expression of invariants of G.*

Sufficient

- Sophie successfully produces sufficient conditions for some specified property.
  - If *expression1 of invariants of G* ≥ *expression2 of invariants of G*, then *G* has property *P*.

reference

Described in …

Hamiltonian Path

- A graph is said to have a Hamiltonian Path if each of its vertices can be visited exactly once in some tour of the vertices.

Evaluating Conjectures

Chvatal & Erdös

G.pc: If $\alpha(G) - 1 \leq \kappa(G)$, then G has H-path.

G.pc: b = 2rad, then G has H-path. new

G.pc: If $n - \sigma - 1 \leq$ a median degree, then G has H-path. Corollary to a result of Chvatal

In all when I stopped the program, there were 30 conjectures on the list.

So far, 7 have been proven, 8 refuted, 1 previously known, and 1 follows from a known result.

reference

`http://cms.uhd.edu/faculty/delavinae/research/wowii`

End of Presentation