Alabama Journal of Mathematics
Spring/Fall 2009

Graph Coloring: History, results and open problems

Vitaly I. Voloshin
Troy University, Troy, AL

Invited Lewis-Parker lecture at the annual meeting of AACTM; Jacksonville State University; Jacksonville, AL; February 28, 2009

Coloring theory started with the problem of coloring the countries of a map in such a way that no two countries that have a common border receive the same color. If we denote the countries by points in the plane and connect each pair of points that correspond to countries with a common border by a curve, we obtain a planar graph. The celebrated Four Color Problem asks if every planar graph can be colored with 4 colors. It seems to have been mentioned for the first time in writing in an 1852 letter from A. De Morgan to W.R. Hamilton. Nobody thought at that time that it was the beginning of a new theory. The first "proof" was given by Kempe in 1879. It stood for more than 10 years until Heawood in 1890 found a mistake. Heawood proved that five colors are enough to color any map. The Four Color Problem became one of the most difficult problems in Graph Theory. Besides colorings it stimulated many other areas of graph theory. Generally, coloring theory is the theory about conflicts: adjacent vertices in a graph always must have distinct colors, i.e. they are in a permanent conflict. If we have a "good" coloring, then we respect all the conflicts. If we have a "bad" coloring, then we have a pair of adjacent vertices colored with the same color. This looks like having a geographic map where some two countries having common border are colored with the same color. Graphs are used to depict "what is in conflict with what", and colors are used to denote the state of a vertex. So, more precisely, coloring theory is the theory of "partitioning the sets having internal unreconcilable conflicts" because we will only count "good" colorings.

The most famous erroneous proof was very instructive. On one hand, it shows an excellent example how drawings may be used in Graph Theory proofs. On the other hand, it explicitly exhibits the limits of drawings. Since the result was considered proven until 1890, it also demonstrated the fact that mathematicians of those times (like nowadays) liked writing papers MORE than reading them. But on the top of all that, nobody could even imagine what a dramatic history was ahead. Formulated first in 1852 by Francis Guthrie (a student (!) of de Morgan), "proved" in 1879 by Kempe, refuted in 1890 by Heawod, the Four Color Problem became one of the most famous problems in Discrete Mathematics in 20th century before in 1977 it became the Four Color Theorem by Appel, Haken, and Koch.

Besides many erroneous proofs (not only that by Kempe), it generated many new directions in Graph Theory. For example, only one sub-direction of chromatic polynomials introduced by Birkhoff in 1912 with the aim to solve the problem by algebraic methods counts more than 500 research papers.

The main idea of the final proof is quite simple - by induction on the number of vertices; but the number of cases is huge. Though the Kempe's proof was erroneous, his idea of alternating paths and further re-coloring of the respective subgraphs was used in the final proof. A path on which two colors alternate is called a Kempe chain. In a plane triangulation, a configuration is a derived subgraph with respect to a separating cycle; it is a subgraph induced by the cycle and all the vertices which are inside the cycle in the plane.

There were the following two basic steps in the proof:

1. Proof that any plane triangulation contains a configuration from a list of unavoidable configurations.

2. Proof that each unavoidable configuration is reducible.

In 1976, Appel, Haken, and Koch, using 1200 hours of computer time, found 1936 (!) unavoidable configurations and proved that all they are reducible. Historically, it was the first time when a famous mathematical problem was solved by extensive use of computers. The final accord in this one-century drama occured when the result was widely announced, the paper was published in 1977, and the University of Illinois even announced it by postage meter stamp of "four colors suffice", a few errors were found in the original proof(!).

Fortunately for the authors, they have been fixed. For the first time, regardless the fact that there is no human being who would check the entire proof (because it contains steps that most likely can never be verified by humans), the problem is considered completely solved.

In 1996, Robertson, Sanders, Seymour and Thomas improved the proof by finding the set of only (!) 633 reducible configurations. Computers, Kempe chains, and some other techniques were used in both proofs. The three consecutive theorems - Theorem about six colors, Theorem about five colors and Theorem about four colors show the main feature of graph coloring: there is a very simple proof for six colors, a relatively simple proof for five colors and an incredible difficult and complex proof for four colors. Thus, the Four Color Problem, formulated by a student, was first "solved" by a lawyer, and really solved with many contributions of the very prominent mathematicians of the century.

Let us imagine year 1912. Just 63 years passed since in 1849 Gauss published his second proof of the fundamental theorem of algebra, saying that every polynomial of degree n with complex coefficients has precisely n roots. Though the final point in the proof was not made until 1920, at that time it appeared that the theory of polynomials is so powerful and universal that it can solve almost any problem.

So, in 1912, Birkhoff, working at Princeton, published the following paper: "A determinant formula for the number of ways of coloring a map", Birkhoff (1912). In this paper, he introduced a function denoted by $P(G, \lambda)$ which gives the number of proper colorings of a graph G using the colors from set $\{1, 2, 3, ...\lambda\}$. Since it is a polynomial, it was called the chromatic polynomial of a graph. Suppose the chromatic number of a graph G is $\chi(G)$; it means there are no colorings using 1, 2, 3, ... $\chi(G) - 1$ colors, and there is at least one coloring using $\chi(G)$ colors. It implies that $P(G, 1) = P(G, 2) = ... = P(G, \chi - 1) = 0$ and $P(G, \chi) \neq 0$. In other words, the numbers λ =1, 2, 3,... $\chi - 1$ all are the roots of $P(G, \lambda)$ and the number $\lambda = \chi(G)$ is not a root. Birkhoff was motivated by the colorings of maps and his basic goal was the following: prove that for every map = planar graph G, $P(G, 4) \neq 0$. That meant that any counter example G to the four color problem must have the root $\lambda = 4$ in its chromatic polynomial. It "only" remained to investigate the behavior of the roots (just one root!) in the chromatic polynomials of planar graphs to solve the famous four color problem.

Birkhoff and others have shown that the chromatic polynomial of a graph on n vertices has degree n, with leading coefficient 1 and constant term 0. Furthermore, the coefficients alternate in sign, and the coefficient of the second term is $-m$, where m is the number of edges. Generally, since Birkhoff, the chromatic polynomial, as function in variable λ was studied in the form

$$P(G, \lambda) = \sum_{i=\chi}^{n} r_i(G)\lambda^{(i)}, \tag{1}$$

where n is the number of vertices, χ is the chromatic number, $r_i(G)$ is the number of feasible partitions induced by colorings with exactly i colors, and $\lambda^{(i)} = (\lambda-1)(\lambda-2)...(\lambda-i+1)$.

Throughout the 20th century, there were many results about the roots of the chromatic polynomials of planar and other classes of graphs but all they miss the main goal: four color problem.

Surprisingly, Birkhoff, Lewis and others, investing so many efforts in study of the roots of chromatic polynomials, did not even characterize the graphs for which all the roots are the integers from the set $\{1, 2, 3, ...\chi - 1\}$ (seems the simplest case!). As sometimes happens in mathematics, new ideas come from "nowhere": in 1955, Benzer discovered the linear structure of DNA molecule; motivated by that Hajnal and Suranyi in 1959 introduced and studied interval graphs which are the subclass of chordal graphs. Later, in 1975, it was discovered that chordal graphs have all the roots from the set $\{1, 2, 3, ...\chi - 1\}$. Dmitriev has found that not only chordal graphs have this property, so the ultimate form of the characterization is the following: a graph G is chordal if and only if for every induced subgraph G' all the roots of the chromatic polynomial are integers from the set $\{1, 2, 3, ...\chi(G') - 1\}$.

The history lesson however, is this: algebraic methods could not help to solve the four color problem but together with combinatorial methods produced many new fruitful scientific directions of research.

Sometimes mathematics, and graph theory in particular, looks like a race for generalizations. In order to generalize graph coloring, Erdos and Hajnal in 1966 have introduced hypergraph colorings: the requirement "adjacent vertices must have different colors" was generalized to "at least two vertices in hyperedge must have different colors". This idea was extremely fruitful and led to many generalizations of graph colorings and many hypergraph classes have been discovered. The special attention was paid to bipartite hypergraphs, normal hypergraphs (related to the weak Berge perfect graph conjecture) and extension of graph coloring to many set systems known long ago, like block designs etc. However, in all such generalizations the basic combinatorial problem was to find the chromatic number of a respective hypergraph, i.e. the minimum number of colors. The problem of finding the maximum number of colors systematically never appeared because for any hypergraph on n vertices the coloring using n colors always existed. From this point of view, the classic coloring theory was the theory for finding the minimum only, i.e. it was evidently asymmetric.

The end to this asymmetry was put in 1993 when the concept of mixed hypergraph was introduced (see the search on Google "mixed hypergraph"). Mixed hypergraph is a triple $H = (X, C, D)$ with vertex set X and two families of subsets, C and D, called C-edges and D-edges respectively. Proper coloring of H is a mapping from X into a set of λ colors in such a way that every C-edge has two vertices of the Common color and every D-edge has two vertices of Different colors. Now the very fundamental problem of colorability appeared: not every mixed hypergraph is colorable. The structure of uncolorable mixed hypergraphs is very general. The first asymptotical investigations created an opinion that coloring theory becomes the theory of uncolorable mixed hypergraphs, and colorability seems to be an island in a darkness...

In a colorable mixed hypergraph, the maximum and minimum number of colors over all proper colorings which use all the colors is called the upper (denoted by $\bar{\chi}(H)$) and lower (denoted by $\chi(H)$) chromatic numbers respectively. Thus, since 1993, the chromatic polynomial of mixed hypergraph was studied in the form:

$$P(H, \lambda) = \sum_{i=\chi}^{\bar{\chi}} r_i(H)\lambda^{(i)}, \tag{2}$$

which evidently and naturally generalizes the formula (1). Formula (2) explicitly shows that in this more general case, the degree of the chromatic polynomial is not n but $\bar{\chi}(H)$, and the leading coefficient is not 1 but $r_{\bar{\chi}}$. When H is just a classic graph or hypergraph, they coincide. Many new additional concepts have been introduced and new results proved, which do and do not have any analogies with the classic colorings. This direction is developing so fast that there are more than 170 papers for 15 years what indicates that coloring theory definitely is taking a new shape.

Perhaps one of the most characteristic features of this new theory are the gaps in the chromatic spectrum discovered in 1998. The fact is that in contrast to the classical coloring theory, there are colorable mixed hypergraphs which do not have any colorings in some number of colors intermediate between χ and $\bar{\chi}$. There are even planar mixed hypergraphs with this property, it was proved that the gap in this case can occur only at 3; the smallest example Kobler and Kundgen (2001) has the chromatic spectrum (0,1,0,1,0,0).

How close and how far is this from the coloring theory of 19th and 20th centuries... For many details, some open problems and the bibliography, reader is invited to visit the regularly updated Mixed Hypergraph Coloring web site, just search by Google for ”mixed hypergraph”.

References

Birkhoff, George, D. (1912). A determinant formula for the number of ways of coloring a map. *Ann. of Math*, 2(14):42–46.

Kobler, D. and Kundgen, A. (2001). Gaps in the chromatic spectrum of face-constrained plane graphs. *Electron. J. Combin.*, 8(1).