One Mathematics

László Lovász

Dept. of Computer Science, Yale University,
New Haven, CT 06510

Mathematics is torn by many division lines.

The most prominent of these runs between Pure and Applied Mathematics. The controversy around Bourbaki focuses on Abstract vs. Concrete. The distinction between Structural Mathematics (whose main results are theorems and proofs) and Algorithmic Mathematics (whose results are algorithms and their analysis) can be traced back to ancient times. There is a deep division (or at least so it appears) between Continuous and Discrete Mathematics.

Some of these divisions are consequences of different working environments: for example, applied mathematicians work under very different financing conditions and success criteria from pure mathematicians. Some others are cultural: various branches of mathematics have their own system of conferences, journals, prizes; a different set of mathematical concepts, basics and perhaps values that are assumed in conversation. And of course the appreciation of the abstract varies very much with personality and individual taste.

These division lines are not completely orthogonal to each other. In fact, often identifications are made like pure–abstract–structural–continuous and applied–concrete–algorithmic–discrete. Sometimes even good–pure–abstract–structural–continuous and bad–applied–concrete–algorithmic–discrete [15] (you may flip some of the coordinates according to taste). But even if we don't have 2^4 or 2^5 different sciences, strong centrifugal forces are obvious.

Should we accept this partition as a fact of life ("I am a bad pure abstract algorithmic discrete mathematician")? Some say that this is good enough, and one should accept the splitting of mathematics into smaller and smaller independent branches as a fact of life.

1

I feel that accepting this would really lead to tragic consequences, and that there is a deep unity of our science that gives it its strength and vitality. I will argue that recent trends in our science make these division lines more complex than they appear; that we have to do our best to bridge these gaps; and that these same new trends may provide means to do so.

1 Three new trends changing the world of mathematics

The size of the community. It is a commonplace that the number of mathematical publications (along with publications in other sciences) has increased exponentially in the last 50 years. Mathematics has outgrown the small and close-knit community of nerds that it used to be. And with increasing size, the profession is becoming more diverse, more structured and more complex.

Mathematicians are conservative people; I don't mean that we are right-wing (as far as I can see, we span the political spectrum just like every other profession), but we don't push for changes: we are reluctant to spend time on anything other than trying to prove that $P \neq NP$ (or the Riemann Hypothesis, or whatever problem keeps us infatuated at the moment). So we pretend that mathematical research is as it used to be. We believe that we find all the information that might be relevant by browsing through the new periodicals on the table in the library, and that if we publish a paper in an established journal, then it will reach all the people whose research might utilize our results.

But of course 3/4 of the relevant periodicals are not on the table there (just in the main section 05 in the AMS subject classification devoted to combinatorics, more than 2500 papers are published each year in at least 10 specialized journals and countless others). And even if one had access to all these journals, and had the time to read all of them, one would only be familiar with the results of a small corner of mathematics.

A larger structure is never just a scaled-up version of the smaller. In larger and more complex animals an increasingly large fraction of the body is devoted to "overhead": the transportation of material and the coordination of the function of various parts. In larger and more complex societies an increasingly large fraction of the resources is devoted to non-productive activities like transportation and information processing. We have to realize and accept that a larger and larger part of our mathematical activity

should, and will, be devoted to communication. This is easy to observe: the number of professional visits, conferences, workshops, research institutes is increasing fast, e-mail is used more and more. The percentage of papers with multiple authors has also jumped. But probably we will reach the point soon where mutual personal contact does not provide sufficient information flow.

There is another consequence of the increase in mass: the inevitable formation of smaller communities, one might say subcultures. These seem to arise on a random basis, but then they persist and determine research directions for quite a long time. One such subculture is Discrete Mathematics – Theory of Computing – Operations Research. I don't see any reason, other than cultural, why computational complexity theory should be embraced by the designers of discrete algorithms, but viewed with grave suspicion by the majority of designers of numerical algorithms.

New areas of application. The traditional area of application of mathematics is physics; and no doubt this area involves the deepest mathematics and the greatest success stories. The branch of mathematics used in these applications is analysis, the real hard core of mathematics.

But in the boom of scientific research in the second half of this century, many other sciences have come to the point where they need serious mathematical tools. Quite often the traditional tools of analysis are not adequate.

For example, biology tries to understand the genetic code: a gigantic task, which is the key to understanding life and, ultimately, ourselves. The genetic code is discrete: simple basic questions like finding matching patterns, or tracing consequences of flipping over substrings, sound more familiar to the graph theorist than to the researcher of differential equations. A question about the information content, redundancy, or stability of the code may sound too vague to a classical mathematician but a theoretical computer scientist will immediately see at least some tools to formalize it (even if to find the answer may be too difficult at the moment).

Even physics has its encounters with unusual discrete mathematical structures: elementary particles, quarks and the like are very combinatorial; understanding basic models in statistical mechanics requires graph theory and probability.

Economics is a heavy user of mathematics — and much of its need is not part of the traditional applied mathematics toolbox. The success of linear programming in economics and operations research depends on conditions of convexity and unlimited divisibility; taking indivisibilities into account

(for example, logical decisions, or individuals) leads to integer programming and other combinatorial optimization models, which are much more difficult to handle.

Finally, there is a completely new area of applied mathematics: computer science. The development of electronic computation provides a vast array of well-formulated, difficult, and important mathematical problems, raised by the study of algorithms, data bases, formal languages, cryptography and computer security, VLSI layout, and much more. Most of these have to do with discrete mathematics, formal logic, and probability.

Which branches of mathematics will be applicable in the near future is utterly unpredictable. Just 25 years ago questions in number theory like how many primes there are between $3 \cdot 10^{200}$ and $4 \cdot 10^{200}$ seemed to belong to the purest, most classical and completely inapplicable mathematics; now related questions belong to the core of mathematical cryptology and computer security.

It would seem that this diversity of applications is another centrifugal force; but I think that, to the contrary, it should strengthen the flow of information across all division lines. No field can retreat into its ivory tower and close its doors to applications; nor can any field claim to be "the" applied mathematics.

New tools: computers. Computers, of course, are not only sources of interesting and novel mathematical problems. They also provide new tools for doing and organizing our research.

There is obviously a large variation in the relationship between mathematicians and computers. Some avoid computers altogether; others are glued to theirs. I use them for e-mail and word processing like most of us; less regularly, I use them for experimentation, and for getting information through the web. I have become addicted to searching in the Mathematical Reviews database, and I find it more and more convenient to get information by browsing through electronic journals and, perhaps even more significantly, through home pages of other mathematicians.

Are these uses of computers just toys or at best matters of convenience? I think not, and that each of these is going to have a profound impact on our science.

It is easiest to see this about experimentation with Maple, Mathematica, Matlab, or your own programs. These programs open for us a range of observations and experiments which had been inaccessible before the computer

4

age, and which provide new data and reveal new phenomena. [1]

Electronic journals and databases, home pages, and e-mail provide new ways of dissemination of results and ideas. In a sense, they reinforce the increase in the volume of research: not only are there increasingly more people doing research, but an increasingly large fraction of this information is available at our fingertips (and often increasingly loudly and aggressively: the etiquette of e-mail is far from solid). But we can also use them as ways of coping with the information explosion.

At first sight, word processing just looks like a convenient way of writing papers. The final output of mathematical research is still a printed paper, read by others in a journal or perhaps more and more from a manuscript printed out in their office. But slowly many features of electronic versions become available that are superior to the usual printed papers: hyperlinks, colored figures and illustrations, animations and the like. A mathematical paper is almost never read in a strictly linear fashion: one jumps back to refresh a definition, jumps ahead to see how a certain lemma is applied, skips proofs at first reading, returns repeatedly to check how the arguments work on a certain example — this is more reminiscent of browsing the internet than of reading a novel. And if a document is not read in a linear way, why write it in a linear way?

I will not discuss here the opportunities (and traps) provided by these features of electronic publication; but it is quite probable that they will gradually transform the way we write papers; and possibly through this also the way we do research.

2 New forms of mathematical activity

The traditional 2500 year old paradigm of mathematical research is defining notions, stating theorems and proving them. Perhaps less recognized, but almost this old, is algorithm design (think of the Euclidean Algorithm or Newton's Method). While different, these two ways of doing mathematics are strongly interconnected (see [17]). It is also obvious that computers have increased the visibility and respectability of algorithm design substantially.

However, as a consequence of the increase in the size of the research com-

---

[1] I don't include computer-assisted proofs, like the proof of the 4-Color Theorem here; I don't feel that they are any different from classical mathematical proofs. The most serious objection to them, namely that few people have the resources to verify them, is obviously not valid any more.

munity, this paradigm must be enriched by new forms of scientific achievement. These may include writing good expositions and surveys, formulating problems and conjectures, compiling examples, experimenting and reporting the results. Let me comment on the first two of these.

Surveys. The most serious threat to the unity of mathematics is the sheer size of mathematical research. No one can read even a tiny fraction of new research papers.

One solution to this problem is the creation of an activity that deals with the secondary processing of research results. For lack of a better word, I'll call this expository writing, although I'd like to consider it more as a form of mathematical research than as a form of writing: finding the ramifications of a result, its connections with results in other fields, explaining, perhaps translating it for people coming from a different subculture.

The community has invented this activity already: there is more and more demand for expositions, surveys, minicourses, handbooks and encyclopedias. Many conferences are mostly or exclusively devoted to survey-type talks; publishers much prefer volumes of survey articles to volumes of research papers.

We organize the Congress every 4 years (and many other regional meetings of the same kind). While some mathematicians feel that the Congress is worthless (and it is if you consider it as a big research conference), people in other fields envy us for it. It is a great asset if used for maintaining the unity of our field, as forum for giving surveys, expositions of the most important new results and new areas and methods of applications.

Yet we all feel reluctance towards accepting expository and survey writing as scientific achievement. There is often a reservation about somebody's writing an exposition of somebody else's new result (I personally feel that this activity should be encouraged instead). If, as suggested, writing expositions should become a highly regarded research activity, one has to find ways of evaluating it. How should surveys fit into our picture of achievements, including jobs, promotions, grants?

We know little about the criteria for marking a good mathematical survey. We don't have a good formal criterion marking a good theorem, either. However, there are some reasonably exact necessary conditions (the theorem should be new, non-trivial, and correct), and the community tends to agree on other criteria that are more difficult to formalize like interest and significance.

Can a survey, or part of it, be counted as a mathematical result? In some

rare cases, the answer may be "yes". A survey making the first connection between two seemingly unrelated areas, or first pointing out some novel applications of a particular theorem, may be cited later just as a theorem would be. But I would not put this as the main criterion for surveys.

Let me propose a radical idea: let us evaluate surveys in the way humanities evaluate their achievements. We tend to look down upon these areas as "soft", and believe that (in contrast to our own "hard and exact" science) success is a matter of luck or, worse, good abilities in self-promotion. Clearly this feeling is far from the truth, and humanities have their own ways of recognizing excellence in intellectual achievements. We could only gain by learning how to do this without our more direct criteria. Our science could only gain by adapting more of the methods from the vast treasury of human thought into our own pursuit of knowledge.

Problems and conjectures. In a small community, everybody knows what the main problems are. But in a community of 100,000 people, problems have to be identified and stated in a precise way. Poorly stated problems lead to boring, irrelevant results. This elevates the formulation of *conjectures* to the rank of research results. Conjecturing became an art in the hands of the late Paul Erdős, who formulated more conjectures than perhaps all mathematicians before him put together. He considered his conjectures as part of his mathematical *œuvre* as much as his theorems. One of my most prized memories is the following comment from him: "I never envied a theorem from anybody; but I envy you for this conjecture."

Of course, with conjectures we run into the same difficulty as with surveys: it is difficult to formulate what makes a good conjecture. And indeed, there is a lot of controversy around the style of Erdős's conjectures. It is easy to agree that if a conjecture is good, one expects that its resolution should advance our knowledge substantially. Many mathematicians feel that this is the case when we can clearly see the place of the conjecture, and its probable solution, in the building of mathematics; but there are conjectures so surprising, so utterly inaccessible by current methods, that their resolution *must* bring something new — we just don't know where.

Another source of conjectures is experimental mathematics, made possible by computers. Among the many examples of this, let me mention the most systematic: the graph-theoretic conjecture-generating program GRAF-FITI by Fajtlowicz [10, 11]. I was very doubtful about a computer's raising conjectures until I got fascinated by one of them, which turned out to relate

to a key question in communication complexity theory.[2]

How to do these experiments properly, how to report their results, how to incorporate them in our science: this is a real challenge. I have already suggested that we should learn from the humanities the proper ways of conducting and evaluating expository writing; let me add that we should look at the experimental sciences for answers to these questions.

3   Discrete and continuous

The most intrinsic among the division lines is discrete vs. continuous, because it involves basic structures and methods of our science. In this last section, which necessarily gets a bit more technical, I put forth some examples demonstrating how much we could lose if we let this chasm grow wider, and how much we can gain by building bridges over it.

Infinite to finite. It is perhaps unnecessary to argue that discrete and continuous mathematics complement each other and that each utilizes methods and tools from the other. We use the finite to approximate the infinite. To discretize a complicated continuous structure has always been a basic method — from the definition of the Riemann integral through triangulating a manifold in (say) homology theory to numerically solving a partial differential equation on a grid.

In spite of this, I feel that the status of applications of discrete mathematical methods in continuous mathematics is less than satisfactory. Perhaps combinatorics has not yet reached the depth and power of analysis or algebra. Perhaps part of the reasons is cultural: a discrete mathematicians is more likely to have studied Galois Theory or the Borsuk–Ulam Theorem than a "classical" mathematician, say, Ramsey Theory or the Max-Flow-Min-Cut Theorem.

Finite to infinite. It is a slightly more subtle thought that the infinite is often (or perhaps always?) an approximation of the large finite. Continuous structures are often cleaner, more symmetric, and richer than their discrete counterparts (for example, a planar grid has a much smaller degree of symmetry than the whole euclidean plane). It is a natural and powerful method to study discrete structures by "embedding" them in the continuous world.

---

[2]To be fair, the same conjecture was raised earlier by Van Nuffelen [18], and disproved later by Alon and Seymour [1].

A classical example is the use of generating functions (with a continuous variable) to analyze the structure of a sequence. But there are many other important examples. Methods from algebraic topology have been used to prove purely combinatorial statements (see [4] for a survey).

The leading theme in combinatorial optimization in the '60's and '70's was the application of techniques of linear programming to combinatorics. It is quite easy to formulate the most important combinatorial optimization problems as linear programs with integrality conditions, and it is quite easy to solve these, if we disregard the integrality conditions; the game is to find ways to write up these linear programs in such a way that disregarding integrality conditions is justified.

The power of tools from elsewhere. To support my plea for the unity of mathematics, let me discuss one recent development in the theory of algorithms. My starting example is a simple algorithmic problem in graph theory: given a (finite) graph G, find a partition of its node set into two classes so that the number of edges connecting the two classes is as large as possible. (In spite of its simplicity, this is a rather important problem; see the monograph of Deza and Laurent [7] for a description of its far-reaching connections.)

Unfortunately, this problem is NP-hard. If you are not familiar with this basic notion of complexity theory, it means roughly that there is no efficient (polynomial time) way to find the best partition (at least subject to the hypothesis that $P \neq NP$). We must settle for less: say, to find an approximately optimal partition.

It is very easy to find a partition where at least half of the edges go across; this was first observed by Erdős in the 60's in a different context. Since no partition can pick up more than all edges, this gives an approximate solution that achieves at least 50% of the optimum.

Can we do better? This really innocent question remained unanswered until fairly recently, almost simultaneously, two important results were obtained: Goemans and Williamson [13] gave an efficient approximation algorithm that gets within 13% of the optimum; and building on a series of weaker results, Hastad [14] proved that no efficient (polynomial time) approximation algorithm at all can do better than 6%.

This is a surprisingly small gap in this kind of problem, but from our point of view, it is more important that both results depend on tools that come from quite unexpected places (and also that are applicable to a number of similar problems).

The negative result is one of several lower bounds on approximability of various optima, proved by similar means. The first of these proofs [9] was an application of a result [3] in the theory of *interactive proof systems*, a very interesting but at first sight quite specialized area in complexity theory. Later improvements revealed that the most important mathematical construction in the proof is an *error-correcting code* obtained by algebraic methods.

The algorithm itself depends on another line of previous results based on connections between distant areas. The key step is the use of *semidefinite optimization*, which is an extension of *linear programming*, building heavily on the *spectral theory* of symmetric matrices. Again, this is not an isolated result; the use of semidefinite optimization (combined with *randomized algorithms*) has been very successful in the design of approximation algorithms.

Probability. This brings me to a topic that seems to bridge most of the division lines in mathematics. The importance of probabilistic methods in combinatorics, graph theory, and the theory of algorithms is exploding. Besides their traditional use in Monte-Carlo methods in integration and simulation, randomized algorithms are used for counting, exact and approximate optimization, primality testing, and one could go on.

In non-algorithmic graph theory, the probabilistic method was first introduced in the '50's by Erdős (see [2]). As a method for showing the existence of objects (graphs, or colorings of a given graph, etc.), it is now basic and extremely powerful. Probability enters proofs of theorems whose statement has nothing to do with probability.

The role of probability is certainly not restricted to combinatorics and graph theory: just let me mention sieve methods in prime number theory, or the explanation of turbulence in terms of statistical mechanics [8].

Deeper unity. Probability theory is only one illustration of the unity in mathematics that goes deeper than just using tools from other branches. Many of the basic questions are not *a priori* discrete or continuous in nature — they can be modeled as discrete problems, or continuous problems.

In the last years, I have worked on sampling algorithms (algorithms generating a uniformly distributed random element from a set that is large and often only implicitly described). This question leads to estimating the mixing time of Markov chains (the number of steps before the chain becomes essentially stationary). From the point of view of this application, it is natural to consider finite Markov chains — a computation in a computer is necessarily finite. But in the analysis, it depends on the particular appli-

cation whether one prefers to use a finite, or a general measurable, state space. All the essential (and very interesting) connections that have been discovered hold in both models. In fact, the general mathematical issue is *dispersion*: we might be interested in dispersion of heat in a material, or dispersion of probability during a random walk, or many other related questions. There is always a *Laplacian* operator that describes one step of the dispersion. The speed of the dispersion is governed by the *spectral gap* of the Laplacian; but if information about the spectral gap if not available, one can also relate the dispersion speed to *isoperimetric inequalities* in the state space. To establish isoperimetric inequalities, one most often constructs (explicitly or implicitly) *multicommodity flows*. (I have been mixing language coming from the classical study of the heat equation with that coming from graph theory; this has been intentional. See Chung [5] for an exposition of some of these connections.)

My second example is more vague. I start with probably the most important series of results in graph theory in the last decade or two, the *Graph Minor Theory* developed mainly by Robertson and Seymour. Recall Kuratowski's classical theorem: a graph can be embedded in the plane if and only if it does not contain two specific graphs (the complete graph K_5 and the complete bipartite graph $K_{3,3}$). The notion of containment can be defined here in several different but equivalent ways; let us settle on "containment as a minor", which means the following: H is a minor of G if it can be obtained from G by deleting edges and nodes, and contracting some edges to single nodes. The class of planar graphs (just like the class of graphs embedable in any other fixed surface) is closed under taking of minors. It is clear therefore that this class can be characterized by excluded minors (just list all minor-minimal non-planar graphs). The point in Kuratowski's Theorem is to show that this set of excluded minors consists of two graphs only.

Wagner formulated the daring conjecture in the 30's that *every class of graphs that is closed under taking of minors can be characterized by a finite list of excluded minors.* The central result in the Robertson–Seymour theory is the proof of this conjecture. However, what I would like to comment on is an "auxiliary" result, which describes, in a sense, large graphs not containing a fixed graph H as a minor. Informally, it says that every such graph can be constructed in the following way: take an arbitrarily large graph embedded in a surface with bounded genus; add edges connecting nodes on the same face at bounded distance; add a bounded number of further nodes; and glue together such graphs along bounded sets of nodes in a tree-like fashion.

11

Above, "bounded" means a bound depending on the graph H but nothing else.

In an even rougher way, this result says that every huge graph not containing a fixed minor is a one-dimensional arrangement of two-dimensional pieces.

We can see the beginnings of a "global" theory of graphs emerging here: what does a huge graph look like? What hidden structures can be identified in this seemingly unstructured universe? Probably there is a more general theory, identifying 3-dimensional, 4-dimensional etc. structures in large graphs. But the formidable difficulties in the Robertson–Seymour theory (stretching over 19 papers now; see [19] as the last published piece) warn us that such a theory will not be easy to establish.

Recently I learned about the work of David and Semmes [6], and I could not help noticing some analogy with the Robertson–Seymour theory. They give a decomposition of a "reasonable" metric space into pieces of different dimension on different scales. Is there more to this analogy?

But speaking of "global" graph theory brings other important results to mind. The classical Regularity Lemma of Szemerédi states that every huge graph can be "decomposed" into a bounded number of pieces that look "random" (the number of pieces depends on the error in approximating randomness; an exact statement would take too much preparation again). Recently, a flow of exciting applications of this basic lemma emerged; see [16] for a survey.

Does Szemerédi's lemma have a more general setting? An indication may be the recent work of Frieze and Kannan [12], showing a connection between Szemerédi's Lemma and low-rank approximation of matrices.

There is no natural way to divide mathematics, but serious communication gaps can arise unless we realize that we have to pay for avoiding them: pay not only with organizational effort but also with research time devoted to expository writing and reading those expositions, to popularizing mathematics and to listening to mathematical problems from various areas of applications.

References

[1] N. Alon and P. Seymour, A counterexample to the rank-coloring conjecture, *J. Graph theory* **13** (1989), 523–525.

[2] N. Alon and J. Spencer, *The Probabilistic Method. With an appendix by Paul Erdős*, Wiley, New York, 1992.

[3] L. Babai, L. Fortnow, and C. Lund, Nondeterministic exponential time has two-prover interactive protocols, *Comput. Complexity* **1** (1991) 3–40.

[4] A. Bjorner, Topological methods, in: *Handbook of Combinatorics* (eds. R.L.Graham, L.Lovász, M.Grötschel), Elsevier, Amsterdam, 1995, 1819–1872.

[5] F.R.K. Chung, *Spectral graph theory*, CBMS Reg. Conf. Series 92, Amer. Math. Soc., 1997.

[6] G. David and S. Semmes, *Analysis of and on uniformly rectifiable sets*, Amer. Math. Soc., Providence, RI, 1993.

[7] M. Deza and M. Laurent, *Geometry of cuts and metrics*, Springer, Berlin, 1997.

[8] A.J. Chorin, *Vorticity and turbulence*, Springer, New York, 1994.

[9] U. Feige, S. Goldwasser, L. Lovász, S. Safra and M. Szegedy, Approximating clique is almost NP-complete, *Proc. 32nd Foundations of Computer Science*, IEEE Computer Society Press, Los Alamitos, 1991, 2–12.

[10] S. Fajtlowicz, On conjectures of Graffiti, *Discrete Math.* **72** (1988), 113–118.

[11] S. Fajtlowicz, On conjectures of Graffiti V, in *Graph theory, combinatorics, and algorithms, (Kalamazoo,MI, 1992)*, Wiley, New York, 1995, 367–376.

[12] A. Frieze and R. Kannan, Quick approximation to matrices and applications (preprint)

[13] M. Goemans and D. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. Assoc. Comput. Mach.* **42** (1995), 1115–1145.

[14] J. Hastad, Some optimal non-approximability results, *Proc. 29th Ann. ACM Symp. on Theory of Computing*, ACM Press, New York, 1997, pp. 1–10.

[15] P. Halmos, Applied mathematics is bad mathematics, in *Mathematics tomorrow* (ed. Lynn Arthur Steen), Springer-Verlag, New York, 1981.

[16] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, in: *Combinatorics, Paul Erdős is eighty, Vol. 2*, J. Bolyai Math. Soc., Budapest, 1996, 295–352.

[17] L. Lovász, Algorithmic mathematics: an old aspect with a new emphasis, in: *Proc. 6th International Congress on Math. Education, Budapest, 1988*, J. Bolyai Math. Soc., 1988, 67–78.

[18] C. van Nuffelen, A bound for the chromatic number of a graph, *Amer. Math. Monthly* **83** (1976), 265–266.

[19] Robertson, Neil and Seymour, P. D., Graph minors. XV. Giant steps, *J. Combin. Theory* Ser. B **68** (1996), no. 1, 112–148.