ADVICE TO MID-CAREER RESEARCHERS

We are starting a new series to provide advice to mid-career researchers. There are a number of programs that SIGMOD organizes for researchers at the beginning of their careers (PhD Symposium and the like) and senior people do not (or should not) need much help. There are considerable challenges for those who are about to transition from an early researcher to a more senior role. In academia, these are people who are about to get tenured that comes with starting to think of moving from shorter-term research objectives to longer-term ones. In industrial research, this corresponds to the transition from participating in projects to initiating and leading them. As a community we don't seem to talk about these challenges much. That is the gap this series attempts to fill. We will get the views of senior researchers from diverse backgrounds and diverse geographies. We will continue as long as we find original advice and the views are not repetitions.

M. Tamer Özsu
University of Waterloo

What Becomes a Senior Researcher

H. V. Jagadish, Univ. of Michigan

As we progress through life, our needs, desires, and even objectives, change. That certainly has been the case for me over the years. This article is my attempt at distilling what advice I can for someone making a transition from a junior to a senior researcher in the database field: think Associate Professor or someone 5 to 10 years post PhD.

Quantitative Metrics

The first thing to notice is that metrics lose importance. We are all used to this already in some regards. For example, we know that grades and standardized test scores, which are solid quantitative metrics, matter less once we are into a PhD program: what counts is research output. We understand that this is because the nature of research is different from what tests measure. We tell our PhD students not to worry about grades so much.

Yet, when it comes to measuring research output, we go immediately to counting papers and citations. I am proud (and I believe rightfully so) of having more lifetime papers in VLDB/PVLDB than anyone else. But I also know that this is not an achievement that really matters in terms of evaluating my research output. What counts is "impact." In fact, the Computing Research Association (CRA) has issued a "Best Practices Memo," which are usually highly influential whitepapers, on why quality matters over quantity in research, and suggested how to focus on the former for faculty hiring and promotion (https://cra.org/resources/best-practice-memos/incentivizing-quality-and-impact-evaluating-scholarship-in-hiring-tenure-and-promotion/).

Still, these counts persist. My hypothesis is that it is because numbers are easier to measure and compare. When there are many candidates to consider, quantitative metrics provide a first filter that we like, even knowing that it is inaccurate. As we get more senior in our careers, there are fewer people with whom we compete, and so metrics lose importance. When we consider applicants for graduate school, it would be painful if we had to make decisions without scores and grades: essays and letters do tell us a great deal, but they are also subject to much happenstance and reasonable people will often disagree. Filtering on a few numeric criteria helps us quickly narrow down to a few top candidates from a large pool. In contrast, when an awards committee is considering a nominee for an honor, say Fellow grade in a society, the decision is based almost entirely on the letters.

As a junior researcher, as you are still trying to establish yourself, you probably tried to publish several papers. This is a reasonable thing to do. In theory, one revolutionary paper should be enough to get you tenure. In practice, that feels risky. But now, having established yourself, you do not need to do that anymore.

By the way, I believe the same thing applies to institutions. As weaker institutions try to establish norms and expectations for research, it is critical to have careful definition of what constitutes an "A" venue and to get numbers. But as institutions become stronger, they too should shift to more holistic, and softer, assessments of impact.

Users

OK, so you are working on fewer things and would like for these to be impactful. How do you do this?

Obviously, you want your efforts to be focused on one topic for maximum impact. When I started (as a junior full professor) at Michigan, I bet on XML, and spent several years building a native XML store. XML turned out not to be as important an area as I had thought it would be—in other words, I made a bad bet---but my Timber XML project still gave me significant recognition.

In my own experience, I have found defining the topic in terms of a system, such as Timber, or some larger multi-faceted problem, to be more effective than defining it in terms of a subject area. The latter does work, but not as well. My work for a decade thereafter, database usability, while impactful, has not had the same "impact per paper," in my subjective opinion, because what connects the papers is only the technology driver and not a well-defined problem objective. I have earned a reputation for my extensive work in the sub-area, but a large visible project would have helped me go farther.

One thing I feel strongly about in choice of problem is the importance of a user for the solution. It is not enough that the problem be interesting to the technologist/inventor: it must matter to a user, even if indirectly. This applies not only at the level of the big picture, but even down to each individual paper. If I cannot clearly visualize someone who would want to use the results of my work, I just do not feel like undertaking it. Of course, I have been wrong about this sometimes, for example, because I misunderstood the user's need. I also apply this test to theoretical results: a lower bound is of value if the problem being bounded is one that a user cares about. We may admire the paper for the elegance of the lower bound proof, but mathematical beauty does not make up for lack of motivation.

I spent my early career at AT&T Labs. The user there was clear: the company I worked for. My question was whether some business unit would potentially care about a line of research that I undertook. AT&T, at least at that time, gave us researchers complete freedom to decide what to work on. Our gratitude to the company led us each to choose problems we thought we would be good at solving, and whose solutions we thought would benefit the company. So, I naturally fell into "good habits" in terms of problem definition.

Once I moved to academia, I keenly felt a lack of users I could connect with. I was fortunate to find myself with an opportunity to work in bioinformatics. From those interactions, I found very smart people, not trained in Computer Science, struggling with simple data management tasks. From there sprung my interest in database usability. Over the years, I have worked extensively with colleagues in biomedicine. But this has never become a true collaboration where I understand enough of the biology to contribute to advance the science myself. Rather, my collaborators have been my "users" who have suggested to me, through their struggles, the next problem I should solve. Over the years, I have worked with colleagues in multiple disciplines in exactly this manner.

The reason users matter is that they force us to think of problems holistically. Database management is perhaps the most insular sub-field in CS. If we start with the relational paradigm, and a small number of products that do a great job with it, we become hammers looking for nails. Understanding user needs opens the door to introducing other ideas, from other sub-disciplines in computing, and even from outside CS altogether.

Persistence

Not every bet will pay off. Sometimes it may be good to cut your losses and get out quickly. However, my own belief is that most of us do that too soon. In my own career, there are two occasions I regret giving up too quickly. The first was in 1995. With Mumick and Silberschatz, I published the first paper about a streaming data system in PODS 1995. It was motivated by a real need at AT&T, and made it into multiple products, but found no traction in the database research community. By the time streaming data became popular a decade later, no one remembered our "Chronicle Data

Model." The second time was probabilistic databases. Nierman and I had a paper in VLDB 2002. Here, I was only a few years ahead of the database community's interest in the topic, and wasn't even the first to write about probabilistic data. But I was definitely early, and could have had much greater impact if I hadn't been so quick to move on.

In both these situations above, I think the right thing to do would have been to stay with the topic. If I had defined and solved additional problems, and published more papers, on related questions, I believe I would have built up a body of work that would have been harder to ignore. Indeed, as an instance of me following my own advice, a couple of years ago, Asudeh and I published our first paper on coverage in a data set. Motivated by bias in data-driven decision systems, our concept was to identify holes in the training data. We struggled to get the new problem definition accepted in the community, and barely squeaked in a paper to ICDE. But since then we have two PVLDB papers and a SIGMOD paper, clearly establishing the sub-topic. (And we are not done: we have more ideas in this area that we are still developing.) I don't know if we have been able to change the minds of our critics who thought this was not the right problem to solve; but for sure we have a critical mass of work that cannot as easily be ignored as a single paper.

In short, having chosen your problem carefully, do stick with it for a while, even in the face of complaints. Persistence pays.

Service

Finally, you are lucky to be in a good place in your career and life. This is a good time to think about how you will pay it forward. Service is an expectation, because it is necessary to advance the field as a whole. Your peers do care. Award committees do care, as well. More importantly, your rewards for your service work will not only be professional, they will also by psychic.

The next question is what type of service work you should do. You have, of course, served on program committees and reviewed papers. That service is essential for the field, and is needed irrespective of your seniority: if you submit 10 papers for review each year, you must review 30 papers each year to maintain karmic balance, assuming 3 reviews per paper and no senior co-authors. But reviewing is a basic necessity and is not the service you aspire to.

My thinking is that you should choose a service project deliberately, just like you choose a research project, and then you should stick with it for long enough to make a mark. Ideally, there will be something you are passionate about that will improve the field. Your project is to make something happen along these lines. As an example, there has been a push towards improving diversity and inclusion in our field, as in so many other areas of computing. Several people have stepped forward with ideas, and with enthusiasm, to help us do better. Not all ideas are equally good, and some will lose enthusiasm quickly. But I sincerely hope that a few years from now our community will recognize some people because their work on this issue has made a positive difference.