Fascinating random networks

Inaugural lecture
Prof. Nelly Litvak
April 20, 2018

/ Department of Mathematics and Computer Science

TU/e Technische Universiteit
Eindhoven
University of Technology

Fascinating random networks

Where innovation starts

Inaugural lecture Prof. Nelly Litvak

Fascinating random networks

Presented on April 20, 2018
at Eindhoven University of Technology

Introduction

Ladies and gentlemen,

My research area is algorithms for complex networks. Suppose you do not know what it is. Then you can search for it, for example, in Google. At that very moment hundreds of algorithms start running for you on two large networks: the Internet and the World Wide Web. By the way, no, they are not the same thing.
Internet is a network of computers, or, routers, that are physically connected by wires. Internet is a technology that enables us to transfer digital information from one computer to another.
A programmer and an artist Barret Lyon managed to visualize the Internet in his Opte project [1]. Look at the colorful fireworks – this is Internet. The colors represent different continents, and shiny white lines in the center are intercontinental fiber cables – the backbone of the Internet. This picture is displayed in the Museum of Modern Art in New York City.
The World Wide Web, on the other hand, is not a physical, but a virtual network. It consists of web pages connected by hyperlinks that refer from one page to another. Web pages are simply documents, like your ordinary Word or Excel file. Your webpage is stored on your computer. When I request it by typing its http://address, this page is transferred from your computer to my computer, using the Internet.
In this lecture, I will tell you about algorithms for networks, and how we can use mathematics in order to design, understand and improve these algorithms.

Networks everywhere

A network has two essential elements. First, we have a collection of objects, called nodes, for example, a set of people or web pages. Second, objects are connected by some relationship. For example, we may connect two people if they are friends. Facebook is a gigantic network of on-line friendships.
Networks are all around us. In road networks, intersections are connected by the roads. In our brain, neurons are connected if they fire together. In a food web, a fox is connected to a mouse by the relationship that a fox can eat a mouse.

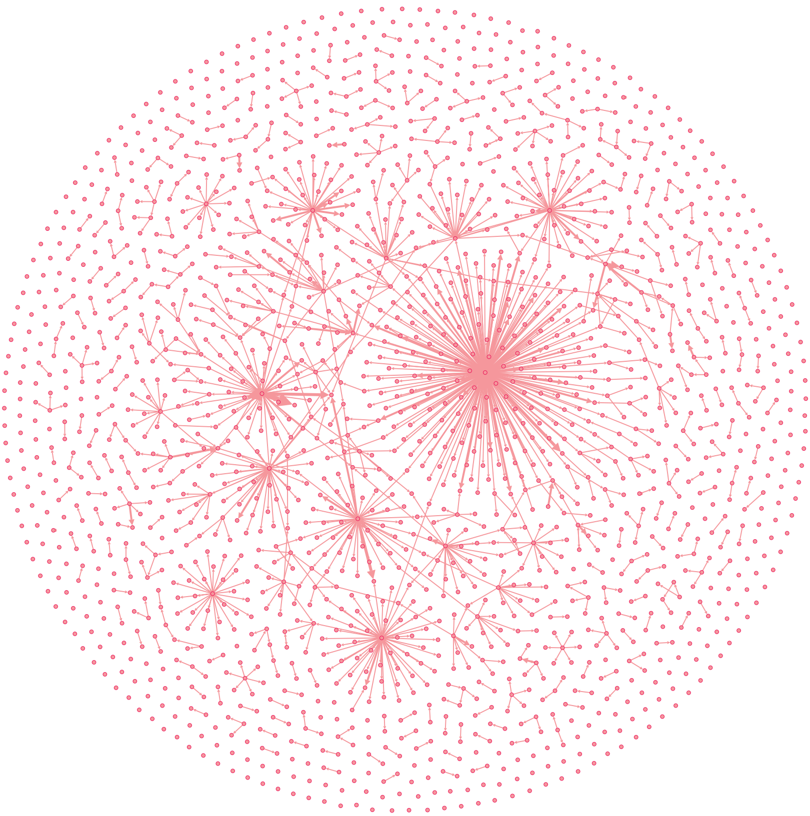

Figure 1

Network of retweets on Project X in Haren, 21-9-2012 07:00. Figure by M.C. ten Thij.

A mathematical representation of a network is very easy. We represent each node as a dot. And if there is a relationship between two dots, we draw a line between them. If the relationship is not symmetric, we denote it by an arrow. For example, I may follow you on Twitter but you might not follow me back. A fox can eat a mouse but a mouse does not eat a fox. A graph is merely a set of dots connected by lines or by arrows.

Figure 1 shows a network of retweets on Project X in Haren in 2012. A birthday invitation of a 16-year-old girl went viral in social media and ended up in a destructive riot. Dots are Twitter users and each tiny arrow represents a retweet from one user to another. This is the network in the morning before the event. Could we have predicted the riot of the same evening by analyzing this network? The interest to these types of questions has been growing quickly in mathematics in the last twenty years. How to find hidden communities? How fast a news item or a virus will spread? What are the network's most influential central nodes? What are the most vulnerable places and connections? Given the scope and importance of open problems, I believe that this research is only in its initial stage.

Science about ideal objects

Mathematical representation of a network as a graph is used in all areas of science. For example, the Human Connectome Project [2] set a goal to build a 'network map' of a human brain. However, there is a difference. In the Connectome project, many measurements and experiments were designed to uncover the relationship between network connections in the brain and human behavior. One can say that the domain knowledge about human brain and behavior was central in this project. But mathematicians focus on the graph itself. They study an idealized abstract object that consists of dots and lines between them. Such abstraction is typical for mathematics, and I would like to say a few words about it.

My daughter is in elementary school and her textbook describes what different sciences are about. 'Biology is a science about living organisms', 'Physics is a science about non-living matter'.

Then I would say, 'Mathematics is a science about ideal objects'. For example, mathematicians study ideal straight lines of zero width and infinite length. No such thing exists in reality. I cannot even draw it because it has zero width. But I am sure you can easily imagine it. And you can also imagine a lot of things that do look like it, for example, a road, a string, or a laser beam. A road is gray and made of asphalt, a laser beam is green and made of light. A mathematician looks at both of them and sees a straight line. By abstracting from the physical nature of the object, mathematics finds commonalities in different systems and provides insights that apply to all of them. This is the power of mathematics.

So, mathematicians create ideal objects and then study their properties. One may ask, if you created an object yourself, you should know all about it, what is there to study? Well, one can also say that we create our children. And then we only hope that we know how they will behave! In fact, mathematicians even use the term *'well-behaved'*, just like parents do; here is an excerpt from Wikipedia:

> Mathematicians (and those in related sciences) very frequently speak of whether a mathematical object – *a function, a set, a space* of one sort or another – is *'well-behaved'*. The term has no fixed formal definition, and *is dependent on context, mathematical interests, fashion, and taste*.
> – From Wikipedia, the free encyclopedia

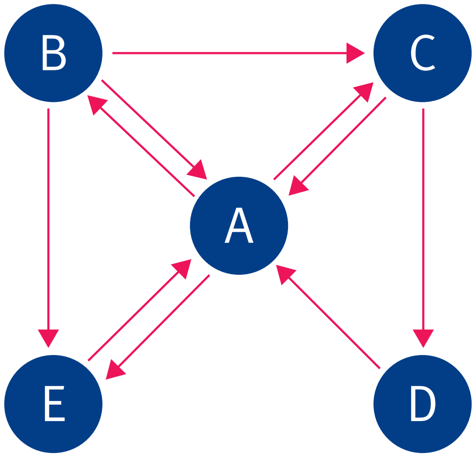

Figure 2

A small directed network.

Once we have created an ideal object, such as a graph, it has a lot of in-built properties that we do not know of and have to discover. Let me show this on a very small directed network in Figure 2. The network has 5 nodes, connected by directed links. It can be, for example, an e-mail network. D wrote to A but A did not write back. We will call the number of incoming links of a node the *in-degree*, and the number of outgoing links the *out-degree*. For example, node A has in-degree 4 and out-degree 3. Let us sum up in-degrees of all nodes, from A to E:

$$4 + 1 + 2 + 1 + 2 = 10.$$

Now, let us sum out-degrees of all nodes, from A to E:

$$3 + 3 + 2 + 1 + 1 = 10.$$

The sums are the same. This is logical because each directed edge has a starting point and a destination. Then, the total number of starting points and destinations must be the same, otherwise some edges will lead to nowhere or come from nowhere. When we construct the network, we do not necessarily aim for these two numbers to be equal, but once we connect dots by arrows, the sums must be the same. It will be true in any directed network of any size. And there is absolutely nothing we can do about it!
Study of ideal objects is very useful because most interesting mathematical objects represent something meaningful in the real world. The graph representation of networks is definitely a prominent example.

Networks and Big Data

The overwhelming interest in networks is very recent; it took off around the end of 1990s. Of course, networks were studied long before that. In 1965, Derek de Solla Price analyzed the network of scientific citations. The papers are the nodes, and a (directed) link means that one paper cites another [9]. It was truly seminal pioneering work. Social networks were studied, too. Milgram conducted his famous small-world experiment in 1960s. The work by Granovetter on the 'strength of weak ties' dates back to 1977. Graphs as mathematical models have a much longer history, introduced already by Leonhard Euler in 1736. Graph theory is a classical branch of mathematics.

So why this sudden massive interest? Today, I started with the Internet and the World Wide Web, the two gigantic networks that were not there before. They gave rise to many other networks: Facebook, Twitter, Wikipedia. Computer technology also allows us to stream, store, analyze and share large amounts of data, including network data. Data has become a crucial game changer in studies of networks. Therefore, I believe that the modern Network Science is an integral part of a broader scope of Data Science.

In mathematics, we often model networks using so-called *random graphs*. In such graphs, the nodes are fixed, but connections are placed at random, according to some probabilistic model. This is a natural approach because links often appear at random, like friendships in a social network.

Also, even if the network is not random, such as the Internet, its structure is so complicated that it is often useful to describe it using statistical summaries and model as a random object.

Random graphs were introduced by Paul Erdös and Alfréd Rényi at the end of 1950s. Initially, they were invented and used to solve difficult problems in graph theory. Notice that this was a very different purpose than modeling World Wide Web, which did not even exist then! The relevance to the network data motivated enormous developments in the theory of random graphs. More and more mathematicians are getting involved in this research. This is just one of the many examples of how a powerful mathematical model becomes useful in the unknown future, even if initially it was designed in a very different context.

Algorithms for complex networks

When I was a little girl, my father, a bio-physicist, played a game with us, called 'A Dummy and a Captain'. Dad was the dummy, and my sister and I were the captains. We had to come up with complicated assignments such as 'Go to the other side of the room, take a pen from the drawer, take a paper from the shelf and write "Hello"'. Then we had to give instructions to the dummy, step by step, one action at a time: 'Stand up, turn left, raise your hand'. Of course, Dad took every chance to mess it up! We would say: 'Take two steps forward', and he would take either two gigantic steps and miss the target or make two microscopic steps and hardly even move. We had to specify the length of the steps.

I recognized this game when I had to write my first computer program. Now the computer was the dummy and took every chance to mess it up! My father was simply teaching us algorithms. An algorithm is a sequence of steps that leads to the result that we want. I will give you some examples of what such algorithms can do on a network, and what challenges they pose.

Sometimes problems are genuinely hard, in the sense that they require a lot of computations even on small networks. For example, you want to distribute a product in a social network, and you have a budget to give this product for free to, say, five people, with the idea that they will promote the product among their friends. Which five people will you choose? This is a very hard problem called *influence maximization* [7].

Sometimes the problem is very easy for a small network, but becomes hard when a network is large or simply is not known to us. For example, how to find the top-100 most followed Twitter users? The Twitter network is not available to us, we can only use so-called *API*. API is an interface that allows one to ask, for example, who are the followers of a specific user. However, the number of requests to API per minute is limited. It will take 900 years to crawl through all Twitter accounts!

In both examples that I gave you it is basically impossible to obtain the exact answers. Randomized algorithms come to the rescue in these situations. The idea is very simple. Forget about getting an exact answer. Let's obtain an approximate answer, but quickly. Such algorithms often involve some kind of random sampling, hence the term 'randomized algorithm'. For example, when my colleagues and I wanted to find the top-100 most followed Twitter users, we sampled 500 random users and used API to see whom they follow. The top-100 followed users will usually be on that list multiple times. Katy Perry has more than 100 million followers. When you sample 500 people at random, it is impossible to miss her! As you see, it is very safe to let randomness decide. Random processes, when repeated many times, behave in a very predictable way.

Designing an efficient randomized algorithm is not always easy, though, because random samples may deviate greatly from average. This is where mathematical modeling is crucial. On a simplified random graph model we can find the mathematical explanation of why the algorithm works or does not work. Then we can predict the performance of the algorithm in different large graphs, and eventually improve it.

Google PageRank

Many qualitative questions on networks have algorithmic solutions. Probably, the most remarkable example is Google PageRank. Google was developed by two graduate students in Stanford, Sergey Brin and Larry Page. In 1998, they published a paper [5], where they introduced the search engine based on entirely new principles. Recently, I told this to the first year Bachelor students, and realized that for most of them 1998 was the year of their birth. They do not know life without Google. Older people, like myself, of course remember that there was life before Google. However, not many realize what an enormous innovation Google was.

Of course, Web search existed before. Maybe some of you remember the search engines like Yahoo! or Alta Vista. What made Google special in the search market? Think about the old, 'non-digital' ways to find information. Before the World Wide Web we could find a phone number in a telephone directory. In such a directory, phone numbers were sorted by topics: schools, doctors, movie theaters. Then, the names were presented in the alphabetic order. A library catalogue was arranged in a similar way. Naturally, when the World Wide Web appeared, people tried to adapt old technologies to the new situation. Yahoo! and Alta Vista were directory-based, attempting to divide web pages into subjects and categories. Google abandoned this idea entirely. Google is 100% query-based. Ask your question, and it will search for you. Of course, Google itself stores information in a very smart way. But the goal of such storage is to simplify the search rather than to divide webpages by subject.

One day, I was driving to Eindhoven from Enschede to give a lecture about the Google PageRank to the students here. On that exact day there was a news item on the radio that the paper version of telephone book would not be printed in the Netherlands anymore. Search-based technology had defeated the directory-based one. In fact, I was thinking: maybe one day everything will be search-based, and we will not even need to learn the alphabetic order?

However, let us go back to PageRank, which was a crucial innovation of Google. It solved the following profound problem. Each query gives you thousands of hits. In which order should we arrange them? For example, if I type *'Dutch Railways'*, I clearly want the official website of the NS, and not the blog of a traveler. But how will a computer know what is important for me?

The revolutionary idea of Google was to use not only the text of the page but also the links to it. This is very logical. If I link to your page, it means I know it and I like it. This is a vote for your page, valuable information, and this was used in the Google search. The PageRank depends on quantity, but also on quality of links to your page. This can be seen on a small example in Figure 3 from the Wikipedia page on PageRank. By the way, Wikipedia by itself is a large network of pages with links to each other, and I often used it for empirical studies in my research. In the figure, the size of the nodes represents their PageRank score. Node B has a large PageRank because it has many incoming links. The PageRank of node C is high because it received the only outgoing link from the important node B. PageRank is a measure of importance of a web page, and we can rank all web pages accordingly. This way, the official website of the NS will receive a very high score, and will be on top of the list for any query related to trains in the Netherlands. PageRank was designed for web search but has been used for many different applications: detecting communities in social networks, combating web spam, or finding most endangered species in food webs. This is because PageRank in fact solves a much broader problem than only ranking the web pages. This problem is known as *network centrality*, and can be formulated as follows: *Given the graph, can we compute which nodes are the most important or central in the network?*

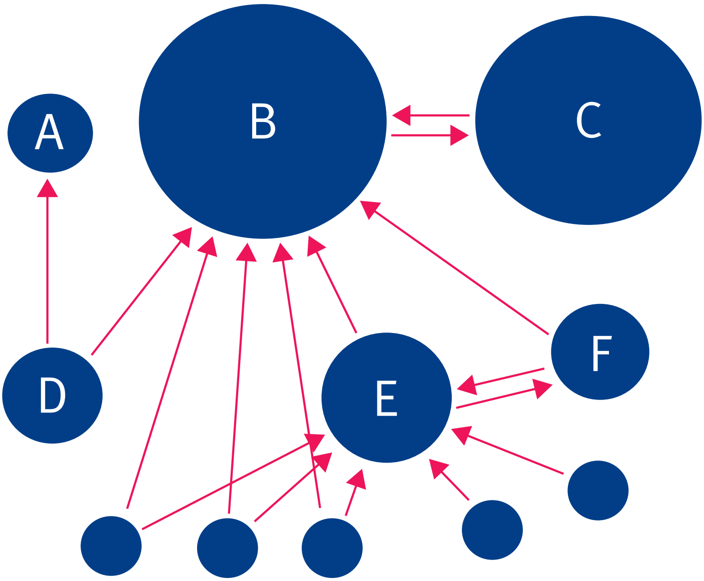

Figure 3

PageRank in a small network. The size of the nodes represents their PageRank score.

Centrality is very important because central nodes can be, for example, crucial hubs in transportation networks or most influential people in social networks. The problem of network centrality is not very new. Bavelas in 1950 [3] and Katz in 1953 [6] already developed the first algorithms to measure influence in social networks, based on the graph only. At the moment, there are many different notions of centrality and many ways to compute it. PageRank is just one of them. It is often convenient and easy to rank nodes by their centralities. However, the properties of such ranking are hard to understand. I will give you an example of one such property for PageRank that is important for robust meaningful ranking in many real-world networks. This is related power laws, which I will first explain.

Power laws

How many connections does a node in a network have? One of the most stunning properties of real-life networks is that the number of connections has huge variability. Some nodes have only few connections, and some have millions. Mathematically, we model this using the so-called 'power law' distributions. I like to compare this to something more familiar, such as the distribution of human length. The length of a human typically follows the well-known normal distribution, as in Figure 4. The values are concentrated around the average, plus-minus small random fluctuations.

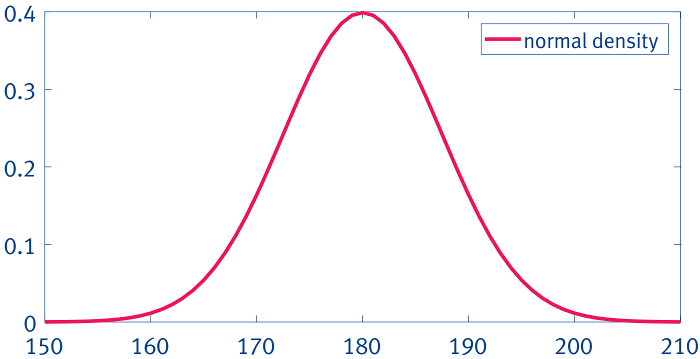

Figure 4

Probability density function of a normal distribution, with average 180 and standard deviation 7.5.

Let's say that the average height of a Dutch man is about 180 cm. Then, if we see a Dutch man, we expect him to be about 180 cm tall, maybe 2 m, but not 10 m. Now, if we take a number of links that point to a web page, the picture is very different. A crawl of the EU web domain in 2015 has 1 070 557 254 pages [4]. The average in-degree is only 85.743 but the maximum in-degree is 20 252 239, which is 236 000 times greater! Human length has 'typical' values, but there is no such thing as a 'typical' web page. This is exactly what the power law captures.

Mathematically, the power law can be written as follows:

$$\frac{\#\text{ nodes with } k \text{ connections}}{\text{total } \#\text{ nodes}} \approx \text{const} \cdot k^{-\tau},\ \tau > 1.$$

Usually, we plot it on the so-called log-log scale, as in Figure 5: on the horizontal axis instead of 1,2,... we plot 1, 10, 100,..., and on the vertical axis we plot 1,0.01,0.0001,....This plot is obviously very different from the normal distribution.

Most of the real-life networks have power laws. This has motivated the development of many new models and tools. I will now show how power laws affect algorithms, such as PageRank, and how we can analyze such phenomena.

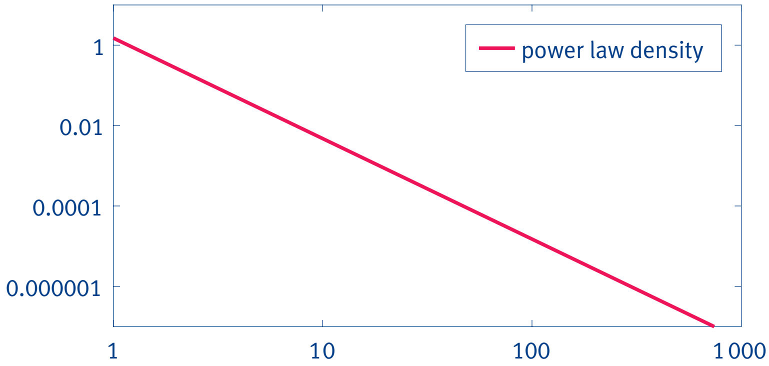

Figure 5

Probability density function of a power law, in the log-log scale, $\tau = 2.5$.

Power law for PageRank

Already from the publication [8] in 2002, we know the following surprising fact. It turns out that if we compute PageRank on a network with a power law degree distribution, then the PageRank will have a power law distribution as well. Figure 6 from our recent paper shows this on a network of citations. We see that on the log-log scale the values of the in-degree and the PageRank roughly follow parallel straight lines. Remarkably, this is true for all scale-free networks in all empirical studies! This is good news, because it means that nodes with the highest PageRank are very different from average; they are stable and easy to identify. Can we *prove* that in a scale-free network PageRank *always* follows a power law? If we could answer questions like this one, we could predict largest PageRank,

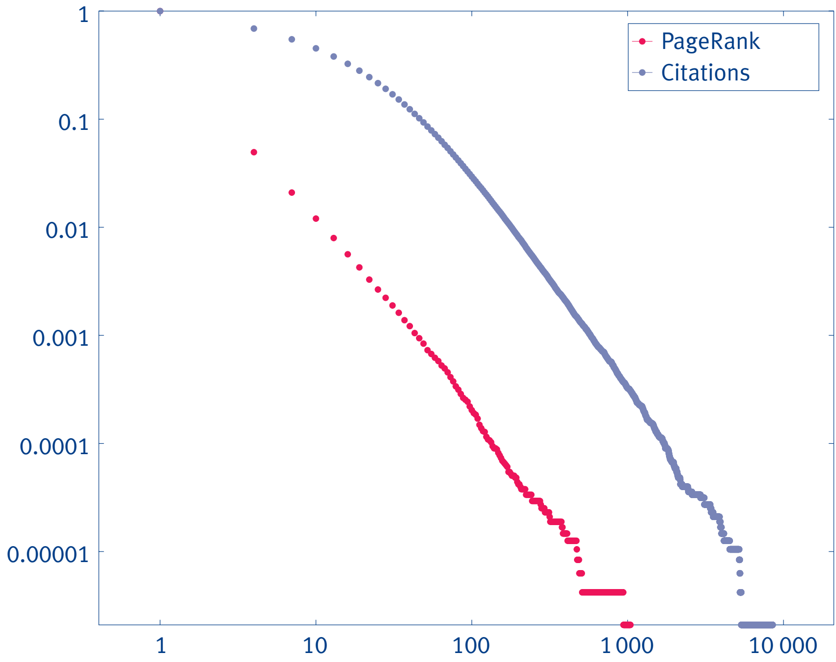

Figure 6

Citation networks from *Web of Science*, Astrophysics. On the *x*-axis are the values of in-degree and PageRank. On the *y*-axis are the frequencies. Blue: in-degree (the number of citations). Red: PageRank. Figure by A. Garavaglia.

investigate its stability, pick up a signal from hidden communities, detect changes and irregularities in the network structure.
Meanwhile, we have developed a systematic approach to the analysis of such problems. In the case of PageRank, we take a random graph, and define the PageRank on this graph. Then we let the number of nodes in the graph go to infinity. The limiting structures are representative for large networks, and they are *well-behaved* because finite-size effects vanish in the limit. With this approach, we now understand the power law behavior of PageRank much better. We have not yet proved the result in full generality, but we are getting there.

Mathematics in the interdisciplinary Network Science

PageRank is just one example, which illustrates how algorithms can be understood and improved, by studying their mathematical properties in random graphs. This research is only in its initial stage. At every step of the analysis, we have to involve new machinery because random graphs are very complicated objects. However, we are steadily gaining more knowledge about them, and it helps. I think in the near future, we can develop a whole set of standard tools for the analysis of algorithms in complex networks. TU/e is a perfect place for this. The Stochastic Section has a strong group on random graphs, funded by the Gravitation Program NETWORKS. I feel grateful and privileged to be a part of this group.

In a broader context, I think we should not forget that the massive developments of the theory of random graphs in the last 20 years have occurred mainly because the data from large networks have become available to us. I like the quote from the Fields Medalist Bill Thurston:

> It's not mathematics that you need to contribute to. It's deeper than that: how might you contribute to humanity, and even deeper, to the well-being of the world, by pursuing mathematics?
> – Bill Thurston

Can we understand a network of bank transactions? Block chains? Optimize Web crawls? Detect trends in the society from Twitter discussions? These are examples that I have recently discussed with companies and researchers from other disciplines.

I have this metaphor for interdisciplinary research. (My daughter, who is an expert in innovation management, said it was 'teenage wisdom' but I will tell it anyway.) I think that interdisciplinary research is like a good marriage. It cannot exist without one of the partners. But it works only if each of the partners stays true to him(her)self. As we all know, a good marriage is not easy. But we have to get out of our comfort zone and talk to each other because most of the real-world problems are too hard to be solved within one discipline only.

I want to create viable collaborations with the Computer Science department here at TU/e. Also, I want to build stronger interdisciplinary ties between TU/e and the

University of Twente, where I have 80% of my appointment. At the UT, I am part of a new Data Science group, with researchers from Computer Science and Electrical Engineering. Also, at the UT, I work with social scientists on on-line social networks. I think all these areas of expertise will be highly relevant for fruitful collaboration, to the benefit of both technical universities.

Teaching to learn

Now, I want to say a few words about teaching. As university teachers, we are privileged because we are the first to meet the new generation of future professionals. We have to be modern, we even need to be ahead of time! We are lucky to live now, because at the moment, education is at the forefront of changes in the world. Almost any information, texts and videos are freely available on-line. I see this as a tremendous opportunity in higher education. We do not need to repeat the books anymore. Instead, we can spend time on interaction with the students. This interaction is one of the best things in our work.
Currently, so much knowledge is being developed that it is not even clear what material we must choose for a modern balanced program. I think it is not a problem if some material is not covered. It is much more important to teach systematic professional approach and problem-solving skills. This can be done in many different ways. I will give you one example.

Recently, in most of my courses, I have implemented a new homework system developed by Eric Mazur of Harvard University. In this system, students are graded not for *correctness* but for *completeness* of their answers. After completing the homework, correctly or not, they discuss it in groups and correct their errors. What I see is that this has taken away the fear of making a mistake. And if you are wondering, the exam results have improved as well.
I like this approach also because I think it is in the spirit of mathematics as science. In mathematics, the reasoning is so much more important than rushing to the correct answer! Mathematics is full of discussions, creativity and making mistakes before we find the right argument. Lewis Carroll was a mathematician, and 'Alice in Wonderland' is full of quotes that I strongly relate to mathematics. Here is one of them: *'Why, sometimes I believe as many as six impossible things before breakfast!'* This could be an accurate description of my working day.
I am excited about new education methods that give more scope for interactions, student independence and creativity. I think this is the best we can do to educate them as future professionals, and I am honored to be a part of it.

Everybody is a math person

I also want to say something about outreach. Unfortunately, mathematics is often perceived as something far removed from real life. This is a major misconception. Today, mathematics does planning and scheduling, protects the data in on-line transactions, and saves lives! One example of the latter is the Kidney Exchange Project for finding optimal swaps between patients and donors. Mathematics is at the core of all digital technologies. The word 'digit' itself means a number. Then why is it that so many people have no idea about the role of mathematics? Recently, I wrote a book 'Who needs mathematics?' about the applications of mathematics in computer technologies. The book became a bestseller in Russia. I gave a lot of public lectures.

Very soon I noticed that the problem was deeper than I thought. People were quickly convinced about applications. But they also wanted to know why they personally needed to learn mathematics. They asked me, 'Why do I need to know what a logarithm is?' Well, why do you need to know where oxygen comes from? Or, what a parliament is? Logarithm is the distance between users in a social network, and the number of kilobytes in your digital picture. At my daughters twelfth birthday, the first thing her friends asked about was not a cake but a WiFi password. Our life increasingly depends on mathematical concepts. Can we really afford *not* knowing what it is?

Last summer, I started a Facebook group called 'Mathematics Great & Terrible'. I try to teach basic mathematics, such as logarithms, to adults with no math background. The group now has about 7000 participants. Unfortunately, what I see is a strong and even painful division between 'math' and 'non-math' people. And both parties believe that there is some kind of 'mathematical gene' that determines whether one can do math or not. This is a very harmful idea that makes it very difficult to reach the 'non-math' audience. And most importantly, recent research shows that this idea is entirely wrong! Stanford professor Jo Boaler is the world's leading specialist in the didactics of mathematics. She says: *'Everybody is a math person'*. Some just need a little bit more time and encouragement.

Many people find mathematical formulas difficult, even scary. I keep telling them that I can do mathematics not because I have some mysterious gene but simply because I was interested and received years of training! Just as in any other

profession. And by the way, this is the only thing I can do for living. Please do not ask me to cut your hair!
As academics we must find a way to explain to each and every person the basic mathematical ideas and what mathematics is about. There are many signs that this will become increasingly important. Popular-science books about mathematics become best-sellers. Recently, Cornell University received a $2.5 million grant to improve math communication, led by the top mathematician and best-selling author Steven Strogatz. Everything confirms the statement that Ionica Smeets made in her inaugural speech:

> Communication of science is not a hobby that a scientist does in free time for a book certificate. It is an essential part of academic work.
> – Ionica Smeets

I am convinced that this work will increase appreciation of mathematics and attract many more talented people to our profession.

Acknowledgments

I want to thank the Executive Board of Eindhoven University of Technology, and Barry Koren, at that time the Interim Dean of the Department of Mathematics and Computer Science, for their trust and support in offering me the professorship position in Algorithms for Complex Networks.

My career started in Nizhny Novgorod, Russia. I want to thank my advisor Mikhail Fedotkin for introducing me to applied probability. At that time research in Russia was in deep economic crisis. My advisor's profound adherence to science supported my motivation to stay in academia.

In 1998 at a conference in Prague I met Willem van Zwet who invited me to apply for position at EURANDOM at TU/e. Dear Willem, thank you for your faith in me. It has changed my career and my life so drastically! I will be always grateful for that.

Ivo Adan was my PhD supervisor here at TU/e. Dear Ivo, working with you was so much fun! Your curiosity for science, as well as your friendly and optimistic attitude were very encouraging for me. I am still very happy about the results we obtained together. My promotor was Jaap Wessels, who is not with us anymore, I have warm memories of him. My second promotor was Henk Zijm from the University of Twente (UT). Dear Henk, thank you for your enthusiasm during my PhD research and later when I joined the UT. I could always feel your support, and it means a lot to me.

Richard Boucherie has been my chair at the UT for 15 years. Dear Richard, we have worked closely together during these years, from co-supervising students to co-chairing the group during your sabbaticals. Thank you for setting a great example in how to see and realize research opportunities. I will keep learning this from you.

I want to thank Maarten van Steen for his guidance in the last years. Dear Maarten, I am very lucky and privileged to have you as a mentor. Thank you for asking the right questions and that you always find the time to talk. I hope I can count on this in the future as well.

In May 2017 I joined the Probability and Statistics group at TU/e part-time. The group is led by Remco van der Hofstad. Dear Remco, I am very happy that you saw the topic of algorithms as a valuable addition to the scope of your group. Thank you for entrusting me with this line of research, this is a dream opportunity for me. I very much enjoy discussing mathematics with you, this is always incredibly illuminating and productive. I look forward to new discussions and new projects.

I feel very much at home at the Stochastic Section at TU/e. Many of the colleagues I know from the time of my PhD, this makes it even more special. I want to thank you all for the warm welcome, and many interesting stimulating discussions.
I want to thank my colleagues at the Stochastic Operations Research group, and the department of Applied Mathematics at the University of Twente, which has been my home for 15 years. Thank you for being around in good and bad times, morning coffees, seminars, research we did together and courses we taught together. A special word of thanks to the department chair Stephan van Gils and Peter Apers, the former dean of the Department of Electrical Engineering, Mathematics and Computer Science. I am grateful that the Executive Board of the University of Twente has approved my appointment as a full professor from January 2018.
I want to thank all students and PhD students that I was fortunate to work with. Especially Yana Volkovich and Pim van der Hoorn who contributed greatly in development of my research line. I am proud of your work and your friendship.
I want to thank my co-authors, especially Konstantin Avrachenkov, Mariana Olvera-Cravioto, Philippe Robert and Dima Shepelyansky. I also want to thank my partners in outreach work, Andrei Raigorodsky and Alla Kechedjan.

I am coming from a very special family of strong, intellectual people, each very successful in his or her own domain. Dear grandpa, grandma, mom, dad, Katya and Peter, you gave me unlimited love and support and enormous intellectual stimulation, throughout my whole life. I have seen you living your dream, failing badly and succeeding spectacularly. From you I have learned that everything is possible. I am greatly indebted to you that I am standing here today.
Dear Natasha, when we moved to the Netherlands 19 years ago, you were my partner in building our new life in the new country. I could count on a six-year-old, and I never forget how special it was. I can still count on you, and I am proud to be the mother of such a strong independent young woman.
Lieve Piyali, je bent een zonnetje in ons huis. Altijd lief en aardig, altijd met een vrolijke lach! En wat een doorzetter ben jij! Ik ben trots op jou en je uitzonderlijke optimisme. Dat moet ik van jou nog leren.
Dear Pranab, thank you for being so kind and patient to me, for your humor, for our travels, and for India in my life. I love that we can talk endlessly with you about mathematics, our students and our courses. I cherish every minute we spend together.

References

1. www.opte.org
2. https://www.neuroscienceblueprint.nih.gov/connectome/
3. Alex Bavelas. Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 1950.
4. Paolo Boldi, Andrea Marino, Massimo Santini, and Sebastiano Vigna. BUbiNG: Massive crawling for the masses. In *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web*, pages 227–228. International World Wide Web Conferences Steering Committee, 2014.
5. Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.
6. Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.
7. David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.
8. Gopal Pandurangan, Prabhakar Raghavan, and Eli Upfal. Using pagerank to characterize web structure. In *International Computing and Combinatorics Conference*, pages 330–339. Springer, 2002.
9. Derek J. De Solla Price. Networks of scientific papers. *Science*, pages 510–515, 1965.

Curriculum Vitae

Prof. Nelly Litvak was appointed part-time professor of Algorithms for Complex Networks in the Department of Mathematics and Computer Science at Eindhoven University of Technology (TU/e) on May 1, 2017.

Nelly Litvak received her MSc degree in Applied Mathematics from the Nizhny Novgorod State University, Russia in 1995 and PhD in Stochastic Operations Research from EURANDOM at Eindhoven University of Technology in 2002. After that she joined the department of Applied Mathematics at University of Twente, where she became full professor in 2018. She was appointed a part-time professor at Eindhoven University of Technology in 2017. Her research is on information extraction and predictions in the large network data, such as on-line social networks and the World Wide Web, randomized algorithms, and random graphs. She is the leader of the 4TU-AMI SRO Big Data, Member of Program Committees for conferences on networks and data mining, and a Managing Editor of the Internet Mathematics journal. She is a best-selling popular science author and gives many public lectures about mathematics and education.

Colophon

Production
Communicatie Expertise
Centrum TU/e

Cover photography
Rob Stork, Eindhoven

Design
Grefo Prepress,
Eindhoven

Print
Drukkerij Snep, Eindhoven

ISBN 978-90-386-4494-3
NUR 919

Digital version:
www.tue.nl/lectures/

Visiting address
Auditorium (gebouw 1)
Groene Loper, Eindhoven
The Netherlands

Navigation address
De Zaale, Eindhoven

Postal address
P.O.Box 513
5600 MB Eindhoven
The Netherlands

Tel. +31 40 247 91 11
www.tue.nl/map

Technische Universiteit
Eindhoven
University of Technology