Snapshots of modern mathematics
from Oberwolfach

№ 10/2019

Limits of graph sequences

Tereza Klimošová

Graphs are simple mathematical structures used to model a wide variety of real-life objects. With the rise of computers, the size of the graphs used for these models has grown enormously. The need to efficiently represent and study properties of extremely large graphs led to the development of the theory of graph limits.

1 Graphs

A *graph* is one of the simplest mathematical structures. It consists of a set of *vertices* (which we depict as points) and *edges* between the pairs of vertices (depicted as line segments); several examples are shown in Figure 1. Many real-life settings can be modeled using graphs: for example, social and computer networks, road and other transport network maps, and the structure of molecules (see Figure 2a and 2b). These models are widely used in computer science, for instance for route planning algorithms or by Google's PageRank algorithm, which generates search results. What these models all have in common is the representation of a set of several objects (the vertices) and relations or connections between pairs of those objects (the edges).

In recent years, with the increasing use of computers in all areas of life, it has become necessary to handle large volumes of data. To represent such data (for example, connections between internet servers) in turn requires graphs with very many vertices and an even larger number of edges. In such situations, traditional algorithms for processing graphs are often impractical, or even impossible, because the computations take too much time and/or computational

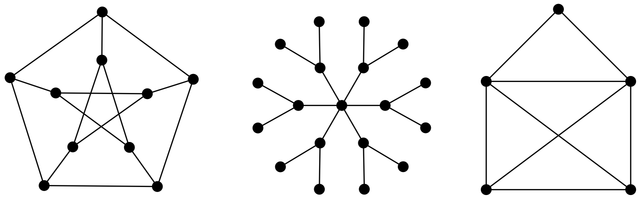
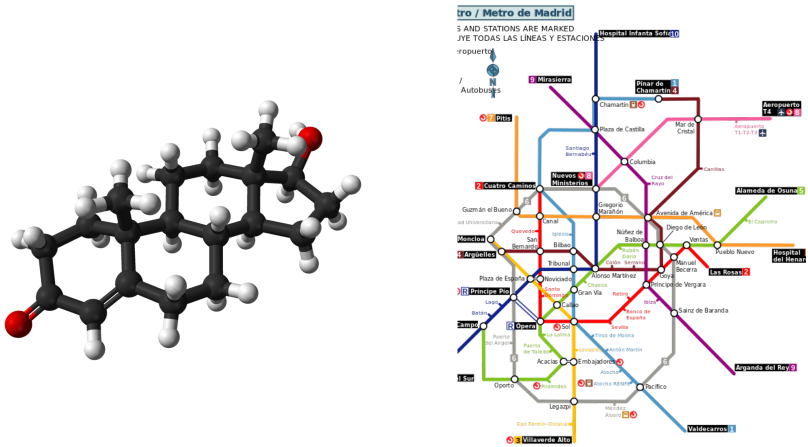


Figure 1: Examples of graphs.

power. This has led to the development of new concepts and methods based on the idea of describing approximate, rather than exact, properties of large graphs. One of these approaches uses the notion of *graph limits*.



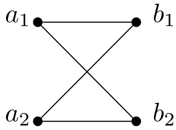
(a) Testosterone molecule.

(b) Metro in Madrid.

Figure 2: Graphs can model molecules or transport networks.

2 Graph limits

Imagine that we have a sequence of graphs, with the property that the number of vertices in each graph of the sequence grows without bound. We want to describe properties of such sequences, and in particular, we want to define what it means for such a sequence of graphs to “converge” to some limit, and describe the limiting object.



	a_1	a_2	b_1	b_2
a_1	0	0	1	1
a_2	0	0	1	1
b_1	1	1	0	0
b_2	1	1	0	0

Figure 3: A *complete bipartite graph* $K_{2,2}$ and its adjacency matrix.

Let us first consider what it means for a sequence of real numbers to converge to a limit. The limit of a convergent sequence of real numbers is itself a real number. Properties of the limit of a sequence often tell us about the sequence itself; for example consider a sequence of numbers that converges to 0. While we cannot say anything about every number in the sequence, we do know that from some moment onwards all elements of the sequence will be "close to 0". The sequence might contain some numbers greater than 1 (or even greater than 1000); however, there can only be a finite number of such terms at the beginning of the sequence.

Our aim now is to define a notion of graph convergence so that we will similarly be able to describe some "limit properties" of a convergent sequence, knowing that all but finitely many graphs in the sequence will be close to having these properties. Unlike in case of convergence of sequence of real numbers, where the limit is again a real number, the limit of a convergent sequence of graphs turns out not to be a graph, but rather an object called a graphon, which is a contraction of "graph function".

Graphons are a relatively young, but intensively studied concept. The notion was first introduced by Lovász and Szegedy in 2006 [4]. They were awarded the 2012 Fulkerson Prize[1] for this innovation. The theory was later extended in a series of papers by Borgs, Chayes, Lovász, Sós and Vesztergombi [1, 2].

3 Graphons

What should the "limit" of a graph look like? To any graph, we can associate an *adjacency matrix*, a matrix consisting of 0s and 1s, with a row and a column for each vertex of the graph. The matrix entry in row v and column w is 1 if there is an edge connecting v to w, and 0 otherwise, see Figure 3 for an example. Note that an adjacency matrix is always symmetric (that is, the entry in i-th row and j-th column is the same as the entry in j-th row and i-th column). A

[1] The Delbert Ray Fulkerson Prize is awarded every three years by the Mathematical Optimization Society and the American Mathematical Society for outstanding papers in the area of discrete mathematics.

$$\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \quad \begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{matrix} \quad \begin{matrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{matrix} \quad \cdots \longrightarrow$$

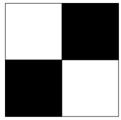


Figure 4: Convergence of adjacency matrices.

graphon is defined to be a symmetric function $W(x, y)$ of two variables, where x and y are in the interval $[0, 1]$, and the output value is also in $[0, 1]$.

Informally speaking, the variables x and y represent vertices and the value $W(x, y)$ determines whether they are connected by an edge. Figure 4 shows a sequence of adjacency matrices and a graphon it converges to: black denotes 1; white denotes 0.

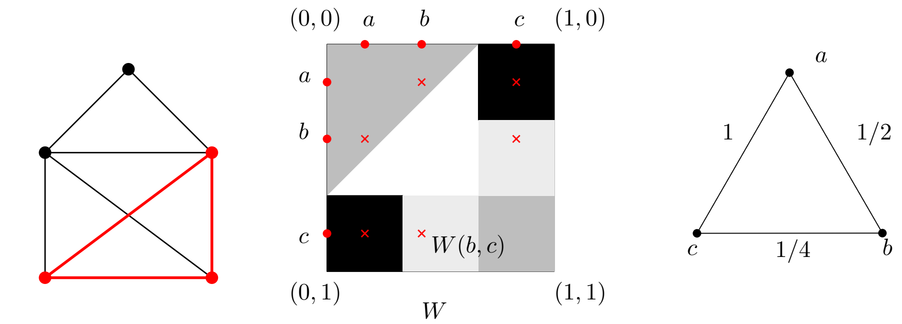
Notice that if we relabel the vertices of a graph, we obtain a different adjacency matrix. In other words, these matrices are not unique. This leads to the question of whether a given sequence of graphs might converge to two different graphon limits. This may actually happen. However, one can define a notion of equivalence between graphons that, roughly speaking, accounts for changes in ordering of the vertices. Equivalent graphons are considered to be “the same”. Taking this into account, every convergent sequence of graphs converges to a unique graphon.

However, note that unlike in the example in Figure 4, graphons are not restricted to have only values zero and one, but any number between 0 and 1. The value $W(x, y)$ of a graphon W at the point (x, y), rather than being a definite 0 or 1 depending on whether or not there is an edge between vertex x and vertex y, should be thought of as the probability that there is such an edge.

To illustrate the idea of the probabilistic presence of an edge, we introduce the notion of a *random graph.* An example of a random graph is a graph with a certain fixed set of vertices where we flip a coin for each pair of vertices, putting an edge between them if we get heads and not putting in an edge if we get tails. Then each edge is present in this random graph with probability $1/2$. If we have a sequence of such random graphs with the number of vertices tending to infinity, we expect that the limit should be a graphon whose value is the constant $1/2$.

How do we know whether a sequence of graphs converges to some graphon? The key to answering this question is the notion of *subgraph density* in a graph. The simplest example of this property is *edge density.* Edge density is the probability that two vertices picked at random are connected by an edge. For instance, the edge density of $K_{2,2}$ (the graph in Figure 3) is $4/6$, because of the 6 possible edges, 4 are actually present in the graph. We can measure edge density in a graphon in a similar way; for instance, the edge density of the graphon in Figure 4 is 1/2, because if we pick two points at random, the value of the graphon on the corresponding coordinates is either 0 or 1, each with probability 1/2.

More generally, we can define the density of a *subgraph* in a larger graph or in a graphon. This is the probability that, if we choose some number of vertices at random, the edges between the vertices we picked form a given pattern, for example a triangle or $K_{2,2}$. See Figure 5 for some examples.



(a) Triangle subgraph in a larger graph.

(b) Probability of a triangle in a graphon: vertices a, b and c form a triangle in the graphon W (where light grey denotes value 1/4 and dark grey 1/2 in the graphon) with probability $W(a,b) \cdot W(b,c) \cdot W(c,a) = 1/2 \cdot 1/4 \cdot 1 = 1/8$

Figure 5: Triangle in a graph and a graphon.

Now we are ready to define what we mean by convergence of a sequence of graphs. A sequence of graphs converges to a graphon if the density of any particular subgraph in the graphs of the sequence converges to the density of this same subgraph in the graphon. This gives us an answer not only to what makes a sequence of graphs convergent, but also to what properties of a sequence are reflected in the properties of the graphon to which it converges.

4 Research on graph limits

The theory of graph limits is still relatively young, yet it has already attracted a lot of research attention and has been used in many areas of mathematics and computer science. In particular, this theory has found many uses in the area of extremal combinatorics, which studies questions of minimizing or maximizing certain quantitative properties of graphs. For instance, the question: "What is the minimum number of triangles a graph with a given edge density can have?", after being studied for many years, was finally answered by Razborov in 2008 [5], using tools closely related to graph limits.

The theory of graph limits has also been used in the area of *parameter testing*, to characterize which parameters of graphs can be estimated by a randomized algorithm that processes only a very small part of the graph [3]. In particular, a parameter f is *testable* if and only if for every convergent graph sequence (G_n), the sequence of numbers $(f(G_n))$ is convergent. For instance, the aforementioned subgraph density is a testable graph parameter for any given subgraph. So, if we want to estimate the density of triangles in a huge graph with a given precision, say 0.01, there is an algorithm which looks only at a small part of our graph (thus saving computational time and space compared to any algorithm which would work with the whole input graph) and returns an estimate of triangle density which, with high probability, say 99%, differs from the actual triangle density in the input graph by less than 0.01. Graphons have found many other applications in probability, statistics and machine learning, where they are used to model large random networks.

As a final remark, let us mention that the theory of graph limits has been so successful for studying graphs that it has inspired the development of theories of limits for other combinatorial structures, including permutations and partially ordered sets.

Acknowledgements

The author is supported by supported by the Czech Science Foundation, grant number 19-04113Y and the Center for Foundations of Modern Computer Science (Charles Univ. project UNCE/SCI/004).

Image credits

Fig. 2a Author: Ben Mills. Public domain, https://commons.wikimedia.org/wiki/File:Testosterone-from-xtal-3D-balls.png

Fig. 2b Author: PerryPlanet. Public domain, https://commons.wikimedia.org/wiki/File:Madrid_metro_map.svg

All other images were created by the author.

References

[1] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Advances in Mathematics **219** (2008), no. 6, 1801–1851.

[2] ______, *Convergent sequences of dense graphs II: Multiway cuts and statistical physics*, Annals of Mathematics **176** (2012), no. 1, 151–219.

[3] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy, and K. Vesztergombi, *Graph limits and parameter testing*, Proceedings of the thirty-eighth annual ACM symposium on Theory of computing, ACM, 2006, pp. 261–270.

[4] L. Lovász and B. Szegedy, *Limits of dense graph sequences*, Journal of Combinatorial Theory, Series B **96** (2006), no. 6, 933–957.

[5] A. A. Razborov, *On the minimal density of triangles in graphs*, Combinatorics, Probability and Computing **17** (2008), no. 04, 603–618.

Tereza Klimošová *is an Assistant Professor at Charles University, Prague.*

Mathematical subjects
Discrete Mathematics and Foundations, Probability Theory and Statistics, Analysis

Connections to other fields
Computer Science

DOI
10.14760/SNAP-2019-010-EN

Snapshots of modern mathematics from Oberwolfach provide exciting insights into current mathematical research. They are written by participants in the scientific program of the Mathematisches Forschungsinstitut Oberwolfach (MFO). The snapshot project is designed to promote the understanding and appreciation of modern mathematics and mathematical research in the interested public worldwide. All snapshots are published in cooperation with the IMAGINARY platform and can be found on www.imaginary.org/snapshots and on www.mfo.de/snapshots.

ISSN 2626-1995

Mathematisches Forschungsinstitut
Oberwolfach gGmbH
Schwarzwaldstr. 9 – 11
77709 Oberwolfach
Germany

Director
Gerhard Huisken

Mathematisches Forschungsinstitut Oberwolfach

Member of Leibniz Leibniz Association

IMAGINARY open mathematics