AKTION 92p1 - Anonymizing User Data: A Parameterized Perspective

The project was mainly carried out by the students Václav Blažej (VB) and Jan Pokorný (JP) at TU Wien. The problem we decided to solve was the Network Microaggregation. The parametrization we choose was by the natural parameters (i.e., the parameters that are "present" on the input) together with structural limitations of the input network (graph). More precisely, we parameterize by the maximum allowed distance to the cluster center (d) and the maximum number of vertices that can form the cluster (u). The structural parameters we derive the complexity for are the treewidth (one of the most prominent and successful structural parameters) and vertex cover (a quite restrictive structural parameter that almost always allows one to design an FPT algorithm).

As the project is now over, we plan to write down the results we got, double-check the proofs, and submit the paper to a conference.

NETWORK MICROAGGREGATION (NMA)

Input: An undirected graph $G = (V, E)$, a length function $\omega\colon E \to \mathbb{N}$, a lower-bound $\ell \in \mathbb{N}$, an upper-bound $u \in \mathbb{N}$, and a maximum allowed distance to a cluster center $d \in \mathbb{N}$.

Question: Is there a partition $\Pi = (C_1, \ldots, C_m)$ of V together with a list of vertices $\mathcal{C} = (c_1, \ldots, c_m)$ such that $\forall i \in [m]\colon \ell \leq |C_i| \leq u, \forall v \in C_i\colon \operatorname{dist}_G(v, c_i) \leq d$.

Week 1 (20.6. - 24.6.) [VB,JP,SS]

In the first week the Czech students (led by Šimon Schierreich) arrived in Vienna and by Tuesday started to get familiar with the new research team/environment. Together with Dr. Ganian and Dr. Simonov they were finding the finer details of the core object of our study (later called Network Microaggregation). More than five variants were formally introduced, and they started to explore first (simple) results.

Moreover, an important part of the first week was to incorporate the students into the research environment within the Algorithms and Complexity Group. Notably, they met the head of the group Prof. Stefan Szeider and other members of the group (Dr. Jiehua Chen,

Week 2 (27.6. - 1.7.) [VB,JP,SS,DK]

The problem definition was fixed on Tuesday and some preliminary results (that transferred from those obtained in the previous week) were presented to the whole team. Then, we proceeded exploring computational complexity of Network Microaggregation (and its connected variant) in more detail. Most importantly, in this week we have derived all the hardness results.

It should be noted that on Wednesday Dr. Knop gave a talk to the whole group about the use of integer linear programming (ILP) in design of algorithms for computational social choice (namely

allocation of indivisible goods). During the talk he presented result using both ILP in fixed and variable dimension. Notably ILP in fixed dimension was later successfully used for some FPT algorithms for the Network Microaggregation problem.

Week 3 (4.7. - 8.7.) [VB,JP]

The aim for this week was on positive results. We believed that by this point we have (nearly) all hardness results we can get and that it should be possible to find some algorithmic results for the remaining cases. It turned out that dynamic programming routines can be designed for some parametrizations when we add a structural parameter treewidth. We believe that this was mainly possible due to Dr. Ganian's presence in the team (as he is a leading expert on the use on dynamic programming on various kinds of structural graph decompositions). Since by the end of the week report the results for bounded treewidth networks were obtained, we have decided that in the next week we should aim on more restrictive parameters---either vertex cover or feedback vertex set.

Week 4 (11.7. - 15.7.) [VB,JP]

We have discovered FPT algorithms for the parameter vertex cover (with other natural parameters of the Network Microaggregation problem). These use integer linear programming and some kernelization (input pruning) techniques. We have discovered some bugs in our former hardness results and managed to fix them during a Zoom call. We decided to set up a shared repository to keep our proof sketches up to date.

Dr. Simonov suggested a couple of well-motivated variants of the Network Microaggregation problem—namely, Connected Clustering on hypercube and Correlation clustering. The students based at TU Wien discovered some hardness results for these problems and (rather) trivial algorithms. Sadly, these variants seem not so much prominent for further deep theoretical study.

Week 5 (18.7. - 22.7.) [VB,JP]

We have finalized the results and decided that our theoretical findings are nontrivial; therefore, we should aim for publication of our results. After the Czech part of the team will be back in Prague, we shall start to write up the proofs. We decided for an online meeting on the 2nd of August. During the meeting we shall discuss the current progress and decide on further plans with the paper we will prepare.

Since in the final week we have found algorithms or (matching) lower bounds for most of the variants of the Network Microaggregation problem, Dr. Ganian introduced a new problem we might aim to solve in future. This problem deals with a road network (represented by an undirected graph) and a set of cars with prespecified trajectories through the graph. The task is to assign to each vertex of the graph and each time slot a traffic light. In our simplified model, we use a type of a junction instead of a type of traffic light. On the positive side, we have some polynomial and FPT algorithms while on the negative side we have a preliminary hardness result even in a very restrictive case. The results we have obtained thus far are not sufficient for publication, however, we plan to pursue this line of research further in a near future (again via, e.g., a Zoom call).