Problem Set 4
CPSC 311 Analysis of Algorithms
Andreas Klappenecker

1. The assignment is due Wednesday, November 1, before class.
2. Email your ranked listed of preferences to me until Tuesday, 5:00pm (sharp deadline).

Estimating Functions. The goal of this project is to investigate methods for estimating functions (such as maxima, minima, cardinality) on sampled data sets. In other words, one tries to infer global properties of a data set by just looking at a small sample. The following quote from reference [M1] nicely motivates this task:

> Estimating the maximum of data sets played a significant role in Allied intelligence efforts during World War II. By analyzing serial numbers from captured German tires, tanks, trucks, flying bombs, and rockets, the Economic Warfare Division of the American Embassy in London, worked to estimate German war production. Beginning in early 1943 with tires, the Economic Warfare division was able to give impressively accurate production estimates. Using a sample, which was later found out to have been 0.3 percent of total production, an estimate was given an error of only 8 percent. ... The estimates for tanks, trucks, bombs, and rockets were also widely successful, and proved to be superior to other means of Allied intelligence gathering for these figures.

Clearly, algorithmic methods for such tasks are very useful and have a wide variety of civil applications.

An interesting problem would be to derive minima and order statistics (the d-th largest element); some solutions are known in the literature. It would be interesting to compare them to the methods used in [M1]. Is it possible to extend the approach taken in [M1] to such functions?

References

[M1] Gum, Lipton, LaPaugh, Fich, *Estimating the Maximum*, J. Algorithms, 54, pages 105-114, (2005).

[M2] Goodman, *Serial Number Analysis*, J. Amer. Statist. Assoc. 47, pages 622–634, (1952).

[M3] See also the references in the above two papers. Feller, volume 1, provides the necessary background from probability theory (see, in particular, page 226).

Expander Graphs. Roughly speaking, an expander graph is a *sparse* regular graph that is at the same time *highly connected*. The two properties are apparently incompatible, so not even the existence of such graphs is obvious. Fortunately, several algebraic and combinatorial constructions are now known.

Expander graphs have numerous applications in computer science. For example, one can derive good families of error-correcting codes from a family of expander graphs. Surprisingly, one can even obtain efficiently decodable codes.

Another striking applications is that one can use expander graphs to derandomize certain randomized algorithms.

It would be interesting to give a survey of expander graph applications in coding theory. An interesting open question is how to obtain self-orthogonal codes from expander graphs.

References

[E1] Davidoff, Sarnak, Valette, *Elementary Number Theory, Group Theory, and Ramanujan Graphs*, Cambridge University Press, 2003
(Google: Valette22-04-02.pdf)

[E2] Expander graphs, course by Linial and Wigderson,
http://www.math.ias.edu/~boaz/ExpanderCourse/

[E3] See also the references given in [E1] and [E2].

Network Coding. Traditional routing in networks is based on the store-and-forward principle. Nodes re-distribute incoming packets to one or more neighbors. It turns out that for many network topologies this is not an optimal routing strategy (particularly in the case of multicasts).

In network coding, the internal nodes are allowed to encode the incoming packets before forwarding them to their neighbors (and a different encoding can be used for each neighbor). One observes that network coding allows one to achieve higher throughput than traditional routing for many network topologies.

Recently, we have been able to determine the expected network coding capacity for mobile ad-hoc wireless networks. It would be interesting to find network coding methods that come close to the network coding capacity for such wireless ad-hoc networks. Distributed randomized network coding solves the problem in the case of static nodes, see [N2]. Apparently, the case of mobile nodes is still an open problem.

References

[N1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network Information Flow", IEEE Transactions on Information Theory, IT-46, pp. 1204-1216, (2000).

[N2] Distributed random network coding for wireless applications http://www.mit.edu/%7Edslun/publications/iwwan2005.pdf

[N3] http://tesla.csl.uiuc.edu/~koetter/NWC/

Problems.

1. Read the introductory sections in [M1]. Describe the algorithm for estimating the maximum given there.

2. Read the introductory chapter in [E1]. State the definition of an expander graph, and of a family of expander graphs.

3. Read [N1]. What is the main result of [N1]?

4. After you have finished problems 1–3, make a ranked list of the three topics according to your preference and e-mail it to me before the deadline.