The Physics of Algorithms: a new approach to information science

PI: M. Chertkov (T-13), Co-Pi: G. Istrate (CCS-DO)
J. Barre (T-11/CNLS, PD), E. Ben-Naim (T-13), L. Gurvits (CCS-3),
A. Hansson (CCS-DO), M. Hastings (T-13), Z. Nussinov (T-11, PD),
A. Percus (CCS-3), E. Ravasz (T-13/CNLS, PD),
M. Stepanov (T-13/CNLS, PD), Z. Toroczkai (T-13/CNLS)

Task:
Analyze existing and design new practice-oriented (i.e. computationally efficient!) algorithms
for the series of inter-related computationally-hard problems of information science:
• Error-correction, storage/reconstruction
• Combinatorial optimization
• Clustering (community detection, data mining) and Coding on complex graphs

New emphasis is on the stat. phys methods of analysis (few hints):
• Efficient (but suboptimal) is (usually) an approximation (saddle-point, Bethe-approximation, replica-
  ansatz mediated approximation, etc) to optimal (but inefficient/expensive)
• error analysis (for quality systems) = rare event analysis
• decoding (instance of the algorithm) -> non-equilibrium stat.phys problem with disorder

data storage
supercomputers
photonics-based, satellite
gene-regulatory

holographic, magneto-optic memory
communications
networks
Error-correction, data storage and reconstruction

Channel Coding

Inference/Reconstruction

Given the detected (real) signal \( x_{out} \)

To find the most probable (integer) pre-image \( \tilde{x}_{in} \)

\[
\text{decoding (ML)} = \max \arg \tilde{x}_{in} \sum_{\tilde{x}_{in} \in \text{allowed}} P(\tilde{x}_{in} | x_{out}) \quad - \text{optimal but inefficient}
\]

Belief Propagation - example of efficient but suboptimal decoding

Algorithms: ANALYSIS + DESIGN

- LDPC, turbo (“random”) codes
- Error floor
- Improving BP
- + inter-symbol-interference
Combinatorial optimization

Satisfiability (SAT)

**Instance**: $N$ variables $x_i = 0, 1$, $i = 1, \ldots, N$ and a Boolean formula composed of $M$ clauses (cnf)

$$F\{x_i\} = \bigwedge_{\ell=1}^{M} C_\ell =$$

$$= (x_1 \lor \bar{x}_7 \lor x_{20}) \land (x_{11} \lor x_3 \lor \bar{x}_{15}) \land \ldots (\bar{x}_9 \lor x_4 \lor \bar{x}_{18}) = \{0, 1\}$$

Problem: “Decide if it exists an assignment that satisfies all clauses simultaneously”

$$F^{-1}(1) \ ?$$

---

Computational Complexity Theory

- **P**: polynomial $t < \mathcal{P}(N)$ (2-SAT, Eulerian circuit, Lin-sys, Primality, ...)
- **NP**: non-deterministic polynomial (checkable in polynomial time) (graph-isomorphism, Compositeness, ...)
- **NP-complete**: hardest in NP (SAT, Coloring, Hamilton-cycle, TSP, Graph-bisection, ...)

---

**Algorithms:**

**ANALYSIS + DESIGN**

- Survey propagation + improvements
- Regular instances
- Source coding (data compression)
Clustering and Coding and Relational Inference on complex Graphs/Networks

Community detection (data mining)

Inference/Reconstruction: determine the most likely distribution of communities.

Optimization: (parallel computing) a process is a node, communicating nodes are connected by edges. Find the allocation (algorithm) which minimizes communication flow.

Example of communities found in an E. Coli genetic network

Coding on graphs

Toy example of network coding: $b_1$ and $b_2$ can be sent/recovered simultaneously

Algorithms: ANALYSIS+DESIGN

- “randomised” vs “regular”
- scalability (efficiency)
- theory for packet-switched simulator
Institutional Impact and Goals
Development of efficient algorithmic solutions to computationally hard problems

Science based prediction of complex systems (Goals 1.3, 1.4):
• “seamless extraction of knowledge from computational datasets that are orders of magnitude larger than generated and exploited currently”
• “develop algorithms and hardware cooperatively”

Preferred laboratory for defense, intelligence and homeland security (Goals 3.3, 4.2):
• communication approaches, information flow and network reliability for large sensor systems
• defensive capabilities in “efficient data communication/operation protocols, and complementary information”

Energy security (Goal 6.2):
• to develop modern communication architectures that “require new approaches such as smart power delivery systems with integrated energy and communication”