Mismatched Monte Carlo for the Planted clique problem

Maria Chiara Angelini

Sapienza, Università di Roma

Rigorous Evidence for Information-Computation Trade-offs Workshop
Simons Institute – EPFL
15 09 2021

The Random Clique Problem

Max-Clique Problem

Given $G \in \mathcal{G}(N, 1/2)$, find the largest random clique

Clique $\mathcal{C}$: fully-connected subgraph of G

The Random Clique Problem

Max-Clique Problem

Given $G \in \mathcal{G}(N, 1/2)$, find the largest random clique

Clique $\mathcal{C}$: fully-connected subgraph of G

$$\overline{Y}_K = \text{Average number of cliques } \mathcal{C} \text{ of size } K = \binom{n}{K} \frac{1}{2}^{\binom{K}{2}}$$
$$\overline{Y}_K \to 0 \text{ if } K > K_S(N) = 2\log_2(N) - O(\log\log(N))$$

The Random Clique Problem

Max-Clique Problem

Given $G \in \mathcal{G}(N, 1/2)$, find the largest random clique

Clique $\mathcal{C}$: fully-connected subgraph of G

$$\overline{Y}_K = \text{Average number of cliques } \mathcal{C} \text{ of size } K = \binom{n}{K} \frac{1}{2}^{\binom{K}{2}}$$
$$\overline{Y}_K \to 0 \text{ if } K > K_S(N) = 2\log_2(N) - O(\log\log(N))$$

Probability that a K-clique is contained in a $K+1$-clique:

$$\mathcal{P}_{grow}(K \to K+1) = \frac{(K+1)\overline{Y}_{K+1}}{\overline{Y}_K}$$

The Random Clique Problem

Max-Clique Problem

Given $G \in \mathcal{G}(N, 1/2)$, find the largest random clique

Clique $\mathcal{C}$: fully-connected subgraph of G

$$\overline{Y}_K = \text{Average number of cliques } \mathcal{C} \text{ of size } K = \binom{n}{K} \frac{1}{2}^{\binom{K}{2}}$$
$$\overline{Y}_K \to 0 \text{ if } K > K_S(N) = 2\log_2(N) - O(\log\log(N))$$

Probability that a K-clique is contained in a $K+1$-clique:

$$\mathcal{P}_{grow}(K \to K+1) = \frac{(K+1)\overline{Y}_{K+1}}{\overline{Y}_K} \to 0 \text{ if } K = (1+\epsilon)\log_2(N)$$

Any polinomial algorithm stops at $K = \log_2(N)$

The Jerrum Metropolis MC

- Start from $x_i = 0 \quad \forall i$
- At each time n choose i u.a.r. and flip x_i with probabilities:

$$P_{\text{Jerrum}}(x_i^n = 0 \rightarrow x_i^{n+1} = 1) = \begin{cases} 0 & \text{if } \exists j : x_j^n = 1 \text{ and } A_{ij} = 0 \\ 1 & \text{otherwise} \end{cases}$$

$$P_{\text{Jerrum}}(x_i^n = 1 \rightarrow x_i^{n+1} = 0) = \lambda^{-1}, \qquad \lambda \geq 1$$

The Jerrum Metropolis MC

- Start from $x_i = 0 \quad \forall i$
- At each time n choose i u.a.r. and flip x_i with probabilities:

$$P_{\text{Jerrum}}(x_i^n = 0 \to x_i^{n+1} = 1) = \begin{cases} 0 & \text{if } \exists j : x_j^n = 1 \text{ and } A_{ij} = 0 \\ 1 & \text{otherwise} \end{cases}$$

$$P_{\text{Jerrum}}(x_i^n = 1 \to x_i^{n+1} = 0) = \lambda^{-1}, \qquad \lambda \geq 1$$

State space: collection Ω of cliques $\mathcal{C}$ of any size in G.

Stationary distribution on Ω: $\pi(\mathcal{C}) = \frac{w(\mathcal{C})}{\sum_{\mathcal{C} \in \Omega} w(\mathcal{C})}$

$w(\mathcal{C}) = \lambda^{|\mathcal{C}|}$: weight asssigned to each clique $\mathcal{C} \in \Omega$

The Jerrum Metropolis MC

- Start from $x_i = 0 \quad \forall i$
- At each time n choose i u.a.r. and flip x_i with probabilities:

$$P_{\text{Jerrum}}(x_i^n = 0 \to x_i^{n+1} = 1) = \begin{cases} 0 & \text{if } \exists j : x_j^n = 1 \text{ and } A_{ij} = 0 \\ 1 & \text{otherwise} \end{cases}$$

$$P_{\text{Jerrum}}(x_i^n = 1 \to x_i^{n+1} = 0) = \lambda^{-1}, \qquad \lambda \geq 1$$

State space: collection Ω of cliques $\mathcal{C}$ of any size in G.

Stationary distribution on Ω: $\pi(\mathcal{C}) = \frac{w(\mathcal{C})}{\sum_{\mathcal{C} \in \Omega} w(\mathcal{C})}$

$w(\mathcal{C}) = \lambda^{|\mathcal{C}|}$: weight asssigned to each clique $\mathcal{C} \in \Omega$

Theorem (Jerrum '92)

Suppose $\epsilon > 0$. For a.e. $G \in \mathcal{G}(N, \frac{1}{2})$ and every $\lambda \geq 1$, the expected time for MC to reach a clique of size at least $(1 + \epsilon) \log N$ exceeds $N^{\Omega(\log N)}$

The Planted Clique Problem

Max-Clique Problem

Given $G \in \mathcal{G}(N, 1/2)$, select u.a.r. a subset $\mathcal{C}$ of size $|\mathcal{C}| \equiv K$.
Add to G all the edges between two nodes in $\mathcal{C}$.
(These operations define the new ensemble $\mathcal{G}(N, 1/2, K)$)
Try to find $\mathcal{C}$.

The Planted Clique Problem

Max-Clique Problem

Given $G \in \mathcal{G}(N, 1/2)$, select u.a.r. a subset $\mathcal{C}$ of size $|\mathcal{C}| \equiv K$.
Add to G all the edges between two nodes in $\mathcal{C}$.
(These operations define the new ensemble $\mathcal{G}(N, 1/2, K)$)
Try to find $\mathcal{C}$.

Possible for $K > 2\log_2 N$

BUT many known algorithms are proved to fail in the regime $K/\sqrt{N} \to 0$:

- Spectral algorithms Alon et al., Random Structures & Algorithms (1998)
- Message Passing Deshpande, Montanari, Found. of Comp. Math. (2015)
- Sum of Squares Barak, SIAM Journal on Computing (2019)

The Jerrum MC for the planted clique problem

Same algorithm as in the random case

Theorem (Jerrum '92)

Suppose $\epsilon > 0$ and $0 < \beta < \frac{1}{2}$. For a.e. $G \in \mathcal{G}(N, \frac{1}{2}, \lceil N^{\beta} \rceil)$ and every $\lambda \geq 1$ the expected time for the MC process to reach a clique of size at least $(1+\epsilon)\log N$ exceeds $N^{\Omega(\log N)}$

But is the MC linear (polynomial) for $\beta \geq \frac{1}{2}$?

A different Metropolis MC for the planted clique

Gamarnik, Zadik, arXiv:1904.07174 (2019)

$x_i = \{0, 1\}$, Fixed global magnetization $m = \sum_{i=1}^{N} x_i \equiv K$.

A different Metropolis MC for the planted clique

Gamarnik, Zadik, arXiv:1904.07174 (2019)

$x_i = \{0, 1\}$, Fixed global magnetization $m = \sum_{i=1}^{N} x_i \equiv K$.

- Choose K nodes $x_{k_i}, i \in [1, K]$ u.a.r. and set $x^0_{k_i} = 1$, $x^0_{j \neq k_i} = 0$

A different Metropolis MC for the planted clique

Gamarnik, Zadik, arXiv:1904.07174 (2019)

$x_i = \{0, 1\}$, Fixed global magnetization $m = \sum_{i=1}^{N} x_i \equiv K$.

- Choose K nodes $x_{k_i}, i \in [1, K]$ u.a.r. and set $x^0_{k_i} = 1$, $x^0_{j \neq k_i} = 0$
- Assign unitary cost for couples of unconnected nodes in the putative clique: $E = \sum_{ij}(1 - A_{ij})x_i x_j$

A different Metropolis MC for the planted clique

Gamarnik, Zadik, arXiv:1904.07174 (2019)

$x_i = \{0, 1\}$, Fixed global magnetization $m = \sum_{i=1}^{N} x_i \equiv K$.

- Choose K nodes $x_{k_i}, i \in [1, K]$ u.a.r. and set $x^0_{k_i} = 1$, $x^0_{j \neq k_i} = 0$
- Assign unitary cost for couples of unconnected nodes in the putative clique: $E = \sum_{ij}(1 - A_{ij}) x_i x_j$
- Update nodes with Metropolis probabilities:

$$P\left((x_a^n, x_b^n) = (1, 0) \rightarrow (x_a^{n+1}, x_b^{n+1}) = (0, 1)\right) = \min(1, e^{-\beta \Delta E})$$

A different Metropolis MC for the planted clique

Gamarnik, Zadik, arXiv:1904.07174 (2019)

$x_i = \{0, 1\}$, Fixed global magnetization $m = \sum_{i=1}^{N} x_i \equiv K$.

- Choose K nodes $x_{k_i}, i \in [1, K]$ u.a.r. and set $x^0_{k_i} = 1$, $x^0_{j \neq k_i} = 0$
- Assign unitary cost for couples of unconnected nodes in the putative clique: $E = \sum_{ij}(1 - A_{ij})x_i x_j$
- Update nodes with Metropolis probabilities:

$$P\left((x_a^n, x_b^n) = (1, 0) \to (x_a^{n+1}, x_b^{n+1}) = (0, 1)\right) = \min(1, e^{-\beta \Delta E})$$

Exponential-in-K time for $K \leq N^{2/3}$

A different Metropolis MC for the planted clique

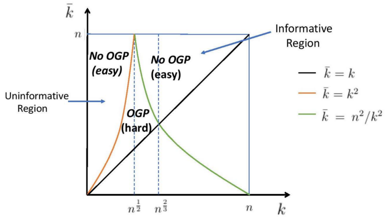
Gamarnik, Zadik, arXiv:1904.07174 (2019)

$x_i = \{0, 1\}$, Fixed global magnetization $m = \sum_{i=1}^{N} x_i \equiv K$.

- Choose K nodes $x_{k_i}, i \in [1, K]$ u.a.r. and set $x^0_{k_i} = 1$, $x^0_{j \neq k_i} = 0$
- Assign unitary cost for couples of unconnected nodes in the putative clique: $E = \sum_{ij}(1 - A_{ij})x_i x_j$
- Update nodes with Metropolis probabilities:

$$P\left((x_a^n, x_b^n) = (1, 0) \rightarrow (x_a^{n+1}, x_b^{n+1}) = (0, 1)\right) = \min(1, e^{-\beta \Delta E})$$

Exponential-in-K time for $K \leq N^{2/3}$



It becomes polynomial down to $K = \sqrt{N}$, working with a mismatched fixed magnetization $\overline{K} > K$.

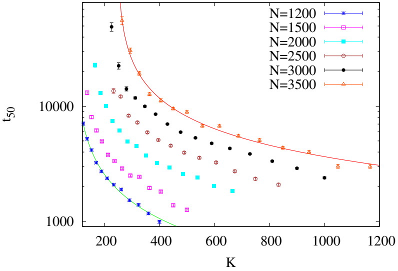
Two questions:

MCA, deFeo, Fachin, arXiv:2106.05720

- Is the Jerrum algorithm suboptimal? (super-polynomial for $K \leq N^{\beta}$ with $\beta > 1/2$)

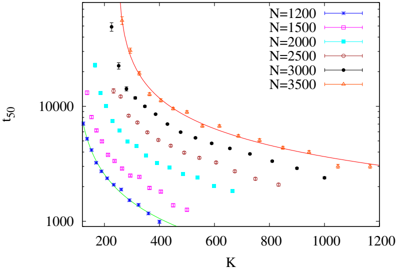
- If yes, can we introduce a mismatched parameter to enhance its performances?

Numerical simulation of Jerrum algorithm

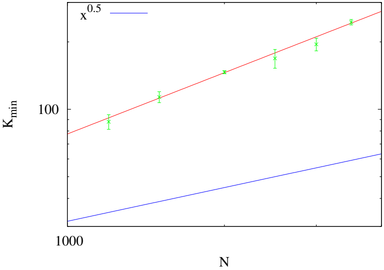


$$t_{50}(K) = \frac{a_N}{(K - K_{min})^{\nu}}$$

Numerical simulation of Jerrum algorithm



$$t_{50}(K) = \frac{a_N}{(K - K_{min})^{\nu}}$$



$$K_{min}(N) = bN^{\alpha}$$

$$\alpha = 0.91$$

Two questions:

- Is the Jerrum algorithm suboptimal? (super-polynomial for $K \leq N^{\alpha}$ with $\alpha > 1/2$)

  Yes, it seems to be suboptimal, $\alpha \simeq 0.91$

Two questions:

- Is the Jerrum algorithm suboptimal? (super-polynomial for $K \leq N^{\alpha}$ with $\alpha > 1/2$)

  Yes, it seems to be suboptimal, $\alpha \simeq 0.91$

- If yes, can we introduce a mismatched parameter to enhance its performances?

  To answer this question we introduce a slightly different MC

BayesMC

Posterior: $P(x|A) = \dfrac{P(A|x)P(x)}{P(A)}$

BayesMC

$$\text{Posterior: } P(x|A) = \frac{P(A|x)P(x)}{P(A)}$$

$$\text{Likelihood: } p(A_{ij} = 1|\{x\}) = \begin{cases} 1 & \text{if } x_i x_j = 1 \\ \frac{1}{2} & \text{otherwise} \end{cases} .$$

$$\text{Prior: } P(x) = \left(\frac{K}{N}\right)^x \left(1 - \frac{K}{N}\right)^{1-x}$$ (local instead of global constraint)

BayesMC

$$\text{Posterior: } P(x|A) = \frac{P(A|x)P(x)}{P(A)}$$

$$\text{Likelihood: } p(A_{ij} = 1|\{x\}) = \begin{cases} 1 & \text{if } x_i x_j = 1 \\ \frac{1}{2} & \text{otherwise} \end{cases}.$$

$$\text{Prior: } P(x) = \left(\frac{K}{N}\right)^x \left(1 - \frac{K}{N}\right)^{1-x} \text{ (local instead of global constraint)}$$

As statistical physicists, we love Gibbs-Boltzmann weights:

$$P_\beta(\{x\}|\{A\}) \equiv P^\beta(\{x\}|\{A\}) \equiv \frac{1}{\mathcal{N}} e^{-\beta H(\{x\})}, \qquad \beta_{\text{Bayes}} = 1$$

introducing the Hamiltonian:

$$H(\{x\}) = -\sum_i \log(P(x_i)) + -\sum_{ij} \left[(1 - A_{ij}) \log \frac{(1 - x_i x_j)}{2} + A_{ij} \log \frac{(1 + x_i x_j)}{2} \right].$$

Metropolis algorithm:

$$P(x_i^n = 0 \to x_i^{n+1} = 1) = \begin{cases} 0 & \text{if } \exists j : x_j^n = 1 \text{ and } A_{ij} = 0 \\ \min\left(e^{-\beta \Delta E}, 1\right) = \min\left(e^{-\beta\left[\log(1-\frac{K}{N}) - \log(\frac{K}{N}) + m\log(\frac{1}{2})\right]}, 1\right) & \text{o.w.} \end{cases}$$

$$P(x_i^n = 1 \to x_i^{n+1} = 0) = \min\left(e^{\beta \Delta E}, 1\right)$$

BayesMC

Metropolis algorithm:

$$P(x_i^n = 0 \rightarrow x_i^{n+1} = 1) = \begin{cases} 0 & \text{if } \exists j : x_j^n = 1 \text{ and } A_{ij} = 0 \\ \min\left(e^{-\beta \Delta E}, 1\right) = \min\left(e^{-\beta\left[\log(1-\frac{K}{N}) - \log(\frac{K}{N}) + m\log(\frac{1}{2})\right]}, 1\right) & \text{o.w.} \end{cases}$$

$$P(x_i^n = 1 \rightarrow x_i^{n+1} = 0) = \min\left(e^{\beta \Delta E}, 1\right)$$

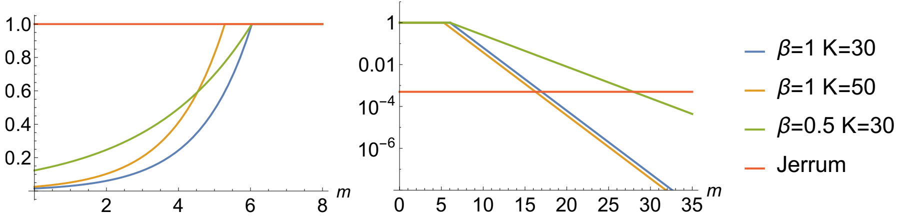
Same class as Jerrum algorithm
Working only on perfect-clique configurations of different sizes m.

Metropolis algorithm:

$$P(x_i^n = 0 \to x_i^{n+1} = 1) = \begin{cases} 0 & \text{if } \exists j : x_j^n = 1 \text{ and } A_{ij} = 0 \\ \min\left(e^{-\beta \Delta E}, 1\right) = \min\left(e^{-\beta\left[\log(1-\frac{K}{N}) - \log(\frac{K}{N}) + m\log(\frac{1}{2})\right]}, 1\right) & \text{o.w.} \end{cases}$$

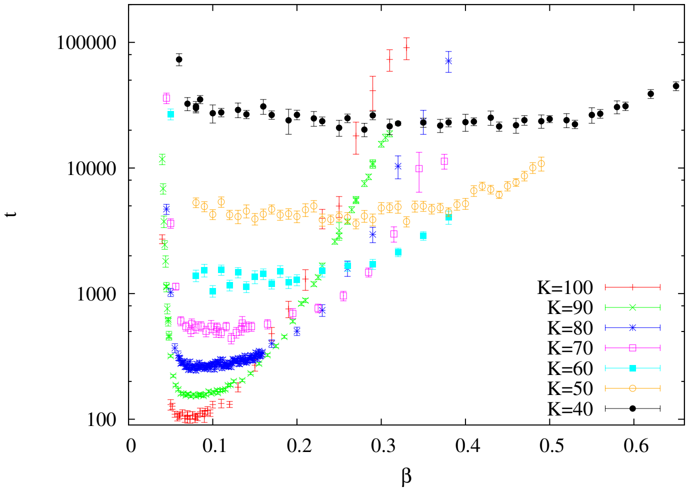
$$P(x_i^n = 1 \to x_i^{n+1} = 0) = \min\left(e^{\beta \Delta E}, 1\right)$$

Same class as Jerrum algorithm
Working only on perfect-clique configurations of different sizes m.



$P(x_i^n = 0 \to x_i^{n+1} = 1)$ $\quad$ $P(x_i^n = 1 \to x_i^{n+1} = 0)$ $\quad$ $N = 2000$

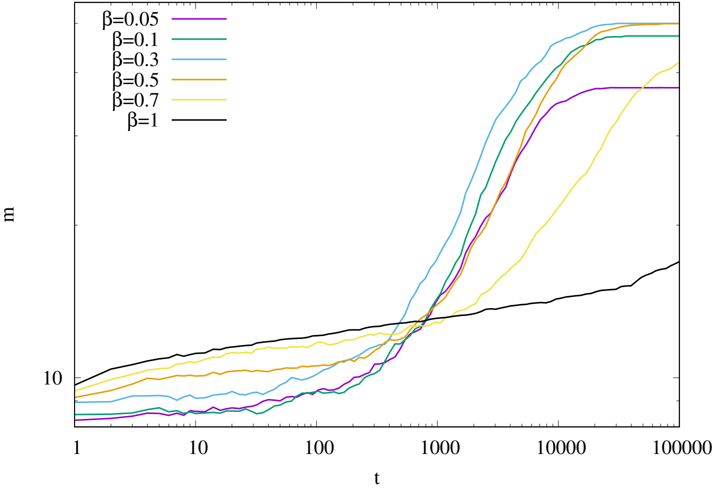
Finding the optimal β



$N = 2000$

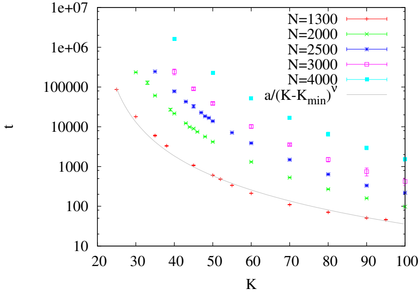
($\beta = 1$: the planted clique is not recovered in $t \leq 10^7$)

Finding the optimal β

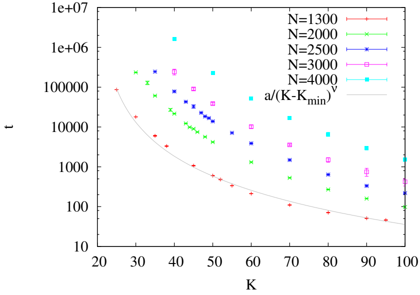


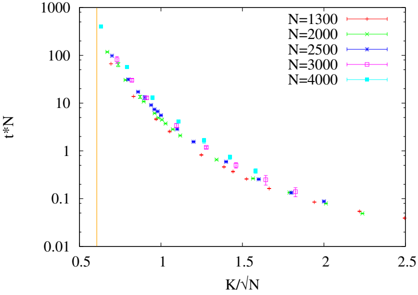
$N = 2000$, $K = 50$

Finding the MC threshold



Finding the MC threshold





Answers to the two questions:

MCA, deFeo, Fachin, arXiv:2106.05720

- Is the Jerrum algorithm suboptimal? (super-polynomial for $K \leq N^{\alpha}$ with $\alpha > 1/2$)

  Yes, it seems to be suboptimal, $\alpha \simeq 0.91$

Answers to the two questions:

MCA, deFeo, Fachin, arXiv:2106.05720

- Is the Jerrum algorithm suboptimal? (super-polynomial for $K \leq N^{\alpha}$ with $\alpha > 1/2$)

  Yes, it seems to be suboptimal, $\alpha \simeq 0.91$

- If yes, can we introduce a mismatched parameter to enhance its performances?

  Yes, we introduce a “temperature”. MC seems to reach the threshold for linear algorithms $K = \sqrt{\frac{N}{e}}$ at “mismatched” temperature $T > 1$.

Conclusions, comments, perspectives

- We look forward for mathematical proofs of our numerical findings

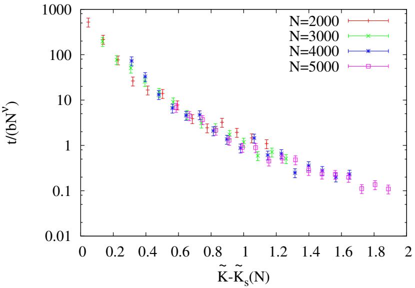
Conclusions, comments, perspectives

- We look forward for mathematical proofs of our numerical findings
- What is the reason for the failure of standard MC? Glassy states/RSB?

Conclusions, comments, perspectives

- We look forward for mathematical proofs of our numerical findings
- What is the reason for the failure of standard MC? Glassy states/RSB?
- Parallel Tempering (*n* exchangeable replicas at different temperatures) works extremingly well for Planted Clique, why?

MCA J. Stat. Mech. (2018) 073404

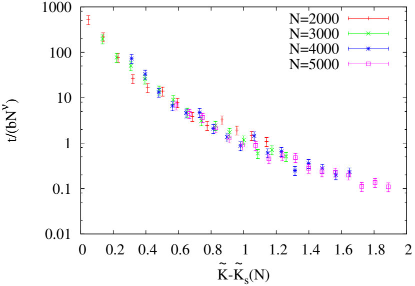


$$\tilde{K} = \frac{K}{\log_2(N)}$$

Conclusions, comments, perspectives

- We look forward for mathematical proofs of our numerical findings
- What is the reason for the failure of standard MC? Glassy states/RSB?
- Parallel Tempering (*n* exchangeable replicas at different temperatures) works extremingly well for Planted Clique, why?

MCA J. Stat. Mech. (2018) 073404



$$\tilde{K} = \frac{K}{\log_2(N)}$$

- Overparametrization is essential in Deep NN. Simple cases can be useful in understanding complex ones.