Learning in Combinatorial Optimization: What and How to Explore

Juan Pablo Vielma

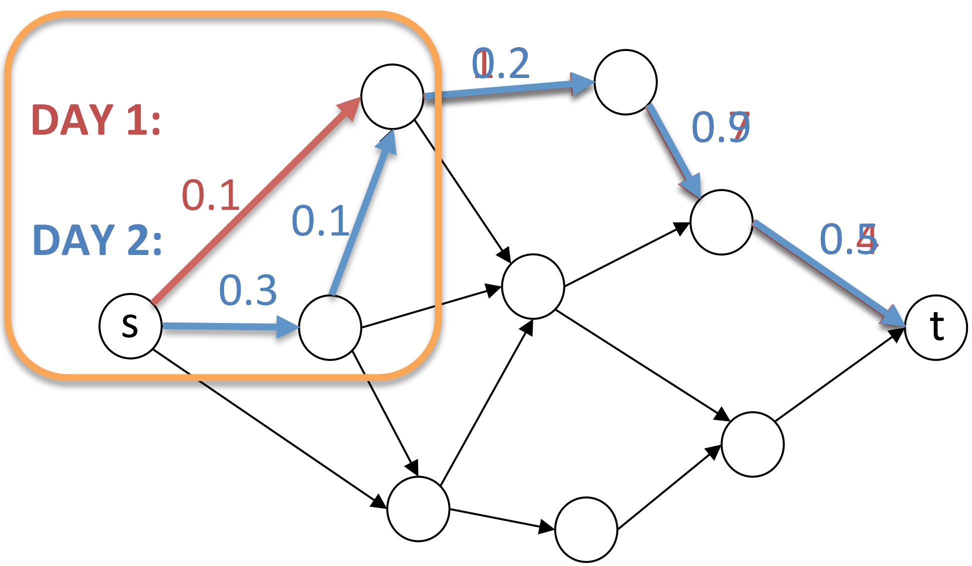
Sloan School of Business, Massachusetts Institute of Technology

Universidad Adolfo Ibañez, Santiago, Chile. October, 2013.

Joint work with D. Saure and S. Modaresi (also J. Orlin and B. Johannes)
Supported by NSF grant CMMI-1233441

Motivation: Driving Home in a New Town

- Shortest ***s-t*** path
- Random edge costs with unknown distribution
- Cost realization observed after usage **(via solution)**



Exploration v/s Exploitation: Bandit Approach

✓ • **What** to exploit: Bandit with best current estimate.

✗ • **What** to explore: All bandits.

✓ • **When** to explore/exploit: Explore with frequency $\frac{\ln N}{N}$

Combinatorial setting: bandits = ***s-t*** paths?

Outline

- Introduction:
  - Problem definition
  - Review of bandit results and direct extensions
- Simple policy = Solution Cover
- Near-optimal policy = Optimality Cover
- Computational Issues
- Simulation Results

Base Problem and Notation

- Base combinatorial optimization problem:

$$f(B) : z^*(B) := \min \left\{ \sum_{a \in S} b_a : S \in \mathcal{S} \right\}$$

feasible solutions (e.g. paths)

$$\mathcal{S} \subseteq \mathcal{P}(A), \quad B = (b_a)_{a \in A} \in \mathbb{R}^A$$

ground sets (e.g. arcs)

costs

- Stochastic version: B distributed according to known F
  - Solve $f\left(\mathbb{E}_F(B)\right)$

Sequential Optimization with On-line Feedback

- Sequence of instances $\{B_n\}_{n=1}^N = \left\{(b_{a,n})_{a\in A}\right\}_{n=1}^N$, for unknown N
- B_n independent, distributed according to initially unknown F
- Only a-priori information on F: $b_{a,n} \geq l_a \quad a.s. \quad \forall a,\, n$
- Need to implement $S_n \in \mathcal{S}$ before B_n is revealed
- B_n is partially revealed after S_n is implemented: $\{b_{a,n} : a \in S_n\}$
- **Goal**: Non-anticipative policy $\pi := (S_n)_{n=1}^{\infty}$:
  - S_n adapted to $\mathcal{F}_n = \sigma(\{b_{a,m} : a \in S_m\,,\, m < n\})$

Performance of Non-anticipative Policy

- Regret relative to clairvoyant agent:

$$R^{\pi}(F,N) := \boxed{\sum_{n=1}^{N} \mathbb{E}_F\left\{\sum_{a\in S_n} b_{a,n}\right\}} - \boxed{N\ z^*\left(\mathbb{E}_F\left\{B_n\right\}\right)}$$

expected cumulative cost of policy

cost of clairvoyant agent

- Expected optimality gap of solution S:

$$\Delta_S^F := \sum_{a\in S} \mathbb{E}_F\left\{b_{a,n}\right\} - z^*\left(\mathbb{E}_F\left\{B_n\right\}\right).$$

- Number of implementations of solution S:

$$T_n(S) := \left|\left\{m < n \ :\ S_m = S\right\}\right|$$

- Alternative form: $R^{\pi}(F,N) = \sum_{S\in\mathcal{S}} \boxed{\Delta_S^F} \boxed{\mathbb{E}_F\left\{T_{N+1}(S)\right\}}.$

independent of policy

policy dependent

Traditional Bandit Approach

- Feasible solutions are singletons: $\mathcal{S} = \{a\}_{a \in A}$
- **Consistent** policies explore all solution with frequency $\ln N / N$

$$\liminf_{N \to \infty} \mathbb{P}_F \left\{ \frac{T_{N+1}(a)}{\ln N} \geq K_a \right\} = 1$$ Lai and Robbins 85

- Optimal regret can be achieved asymptotically

$$\underbrace{\sum_{a \in A} \Delta^F_{\{a\}} K_a}_{\text{Lai and Robbins 85}} \leq \frac{R^\pi(F,N)}{\ln N} \leq \underbrace{\sum_{a \in A} \Delta^F_{\{a\}} \tilde{K}_a + o(1)}_{\text{e.g. Auer et al 02}}$$

  - Optimal Regret is proportional to $|A| \ln N$

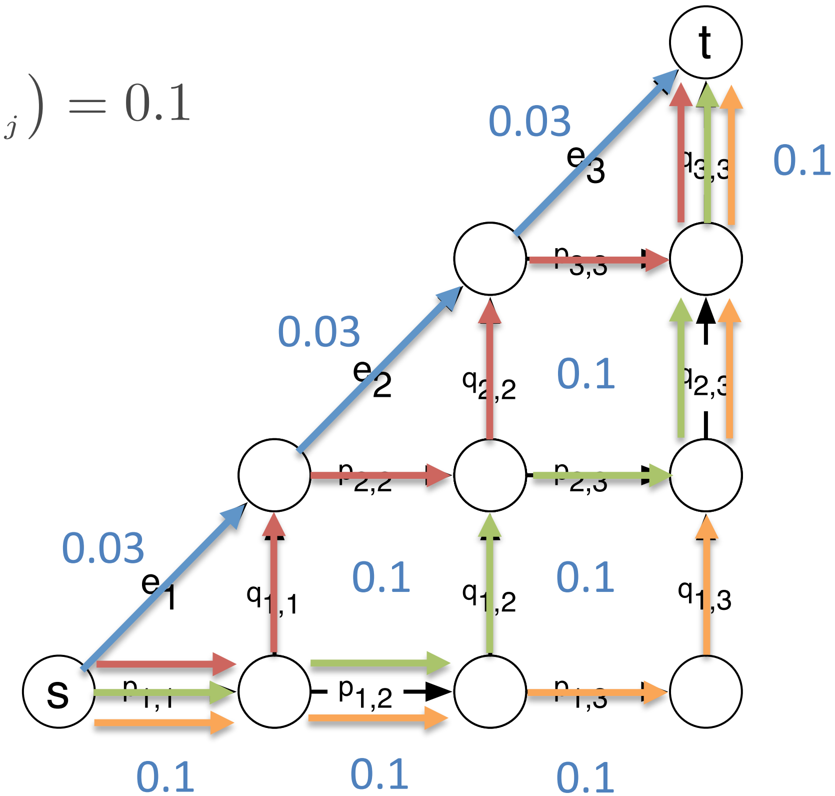
- Naïve adaptation: explore every path with frequency $\ln N / N$?
  - Regret proportional to $|\mathcal{S}| \ln N$!

Performance of Naïve Adaptation

- Instance for k=3 with $l_a = 0$
  - $\mathbb{E}(b_{e_i}) = 0.03, \quad \mathbb{E}(b_{p_{i,j}}) = \mathbb{E}(b_{q_{i,j}}) = 0.1$
  - $|A|$ = (k+2)(k+3)/2
- Optimal cost = 0.03 k
- Non-negative costs = explore all paths
  - Regret = # of paths × $\ln N$:

$$\frac{4^{k+1}}{(k+1)^{3/2}\sqrt{\pi}} \ln N$$

  - Exponential on k and $|A|$!
- Solution: explore all arcs with a path cover of size k+1



A Simple Policy Based on Solution Covers

- **What to Exploit:** Optimal solution to $f\left(\overline{B}_n\right)$ with

$$\bar{b}_{a,n} := \frac{1}{T_n(a)} \sum_{m<n \,:\, a \in S_m} b_{a,m}.$$

- **How to Explore:** Solution cover $\mathcal{E}$ of A

$$\mathcal{E} \subseteq \mathcal{S} \text{ s.t. } A \subseteq \bigcup_{S \in \mathcal{E}} S$$

- **When to Explore:** with frequency $\ln N / N$
  - Cycles with exponentially increasing lengths

Cycles: Exploration Frequency and Performance

- Traditional bandit algorithm of Auer et al 02 = **UCB1**:

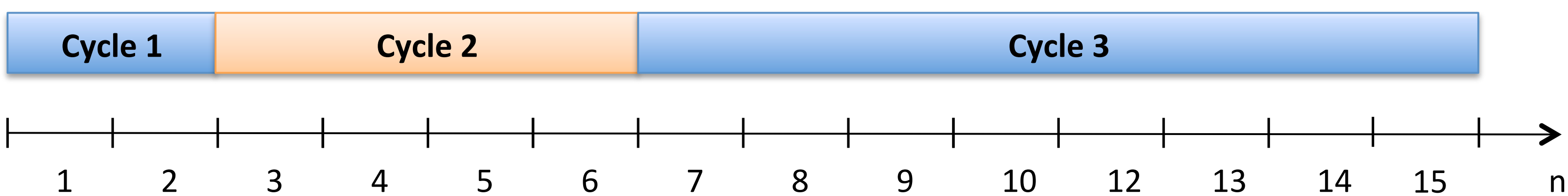
$$S_n \in \underset{S \in \mathcal{S}}{\operatorname{argmin}} \left\{ \bar{b}_{S,n} - \sqrt{2 \ln(n-1)/T_n(S)} \right\} \qquad \bar{b}_{S,n} := \frac{1}{T_n(S)} \sum_{m<n\,:\, S_m = S} \sum_{a \in S} b_{a,m}.$$

current cost estimate

exploration penalty

  – Solves optimization problem every period.

- Exploration/Exploitation cycles with exponential lengths



  – At cycle **i** make sure all arcs have been explored **i** times, then exploit.

Cycle starts $\Phi := \{n_i : i \in \mathbb{N}\}$ with $n_i := \max\left\{\lceil e^{i/H} \rceil, n_{i-1}+1\right\}$

tuning parameter

  – Re-compute optimal solution to $f\left(\bar{B}_n\right)$ only once per cycle.

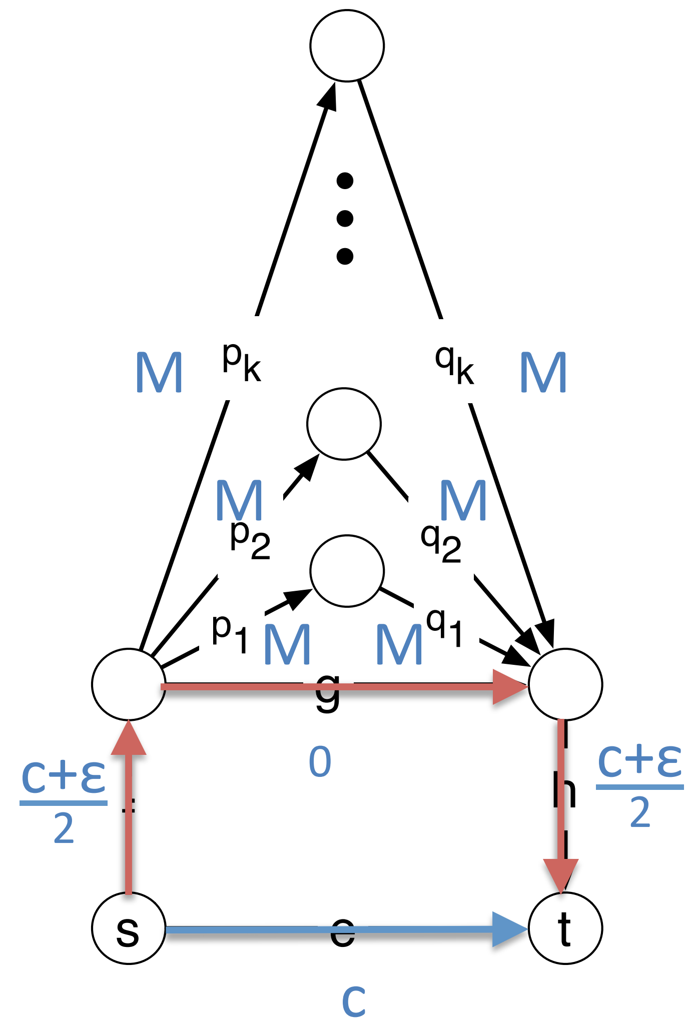
A Simple Policy with regret ≤ $|A| \ln N$

Algorithm 1 Simple policy $\pi_s(\mathcal{E})$

Set $i = 0$, and $\mathcal{E}$ a minimal cover of A
for $n = 1$ to N **do**
update optimum
  if $n \in \Phi$ **then**
    Set $i = i + 1$
    Set $S^* \in \mathcal{S}^*(\overline{B}_n)$
  end if
optimal set
explore
  if $T_n(a) < i$ for some $a \in S$, for some solution $S \in \mathcal{E}$ **then**
    Implement such a solution, i.e., set $S_n = S$
  else
exploit
    Implement $S_n = S^*$
  end if
end for

Are Solution Covers Enough?

- Non-negative costs
- Solutions = k+2 = cover size
- Regret of simple policy with cover is $(kM + \varepsilon) \ln N$
- Regret of simple policy with $\mathcal{E} = \{(f, g, h), (e)\}$ is $\varepsilon \ln N$
- Explore only what is necessary to confirm optimality.



Efficient Exploration = Optimality Cover Problem

$$
\begin{aligned}
OCP(B): \quad \min \quad & \boxed{\sum_{S \in \mathcal{S}} \Delta_S^F(B)\, y_S} \\
s.t. \quad & \boxed{x_a \leq \sum_{S \in \mathcal{S}: a \in S} y_S, \quad a \in A} \\
& \boxed{\sum_{a \in S} \left(l_a (1 - x_a) + b_a x_a \right) \geq z^*(B), \quad S \in \mathcal{S}} \\
& x_a,\, y_S \in \{0, 1\}, \quad a \in A, S \in \mathcal{S},
\end{aligned}
$$

- **What** to explore: arcs needed to guarantee optimality.
- **How** to explore: use a min-regret cover of these arcs.

An Adaptive Policy with "Near-Optimal" Regret

Algorithm 2 Adaptive policy π_a

Set $i = 0$, $C = A$, and $\mathcal{E}$ a minimal cover of A
for $n = 1$ to N **do**
  if $n \in \Phi$ **then**
    Set $i = i + 1$
    Set $S^* \in \mathcal{S}^* \left(\overline{B}_n\right)$
    if $(C, \mathcal{E}) \notin \Gamma(\overline{B}_n)$ **then**
      Set $(C, \mathcal{E}) \in \Gamma^*(\overline{B}_n)$
    end if
  end if
  if $T_n(a) < i$ for some $a \in C$ **then**
    Try such an element, i.e., set $S_n = S$ with $S \in \mathcal{E}$ such that $a \in S$
  else
    Implement $S_n = S^*$
  end if
end for

update exploration & exploitation set

explore

exploit

optimal set of OCP

Implementation: Solving OCP

$$
\begin{aligned}
OCP(B): \quad \min \quad & \sum_{S \in \mathcal{S}} \Delta_S^F(B)\ y_S \\
s.t. \quad & x_a \le \sum_{S \in \mathcal{S}: a \in S} y_S, \quad a \in A \\
& \left\{ \begin{aligned} & \sum_{a \in S} (l_a(1 - x_a) + b_a x_a) \ge z^*(B), \quad S \in \mathcal{S} \\ & x_a,\, y_S \in \{0,1\}, \quad a \in A, S \in \mathcal{S}, \end{aligned} \right.
\end{aligned}
$$

Exponential # of variables and constraints

- Theoretical Complexity of OCP = Bad news?
  - OCP is not guaranteed to be in NP!
  - OCP is in NP when $f(B)$ is in P
  - OCP for **matroids** is in **P**, but for **shortest path** is **NP-hard**

Good News on Solving OCP

- If $f(B)$ has a IP formulation $\{y^S\}_{S\in\mathcal{S}} = \left\{y \in \{0,1\}^{|A|} : My \leq d\right\}$

  then OCP can be "effectively" solved by branch-and-cut.

$$\min \quad \sum_{i=1}^{|A|} \left(\sum_{a\in A} b_a y_a^i - z^*(B) \right)$$

$$s.t. \qquad x_a \leq \sum_{i=1}^{|A|} y_a^i, \quad a \in A$$

Separation by solving $f(B)$

$$My^i \leq d, \qquad i \in \{1,\ldots,|A|\}$$

$$\sum_{a\in S} (l_a(1-x_a) + b_a x_a) \geq z^*(B), \quad S \in \mathcal{S}$$

Polynomial # of variables

$$x_a,\, y_a^i \in \{0,1\}, \quad a \in A, i \in \{1,\ldots,|A|\}.$$

$f(B)$ with LP = OCP with Compact IP

- Example: Shortest path.

$$\min \quad \sum_{i=1}^{|A|} \left(\sum_{a \in A} b_a y_a^i - z^*(B) \right)$$

$$s.t. \qquad x_a \leq \sum_{i=1}^{|A|} y_a^i, \qquad a \in A$$

Feasible Paths

$$\left\{ \sum_{a \in \delta_{out}(v)} y_a^i - \sum_{a \in \delta_{in}(v)} y_a^i = \{0, 1, -1\}, \qquad v \in V,\ i \in \{1, \ldots, |A|\} \right.$$

Optimality with LP duality

$$\left\{ \begin{array}{ll} l_{(u,v)}(1 - x_{(u,v)}) + b_{(u,v)} x_{(u,v)} \geq w_u - w_v, & (u,v) \in A \\ z^*(B) \leq w_s - w_t & \end{array} \right.$$

$$x_a,\ y_a^i \in \{0, 1\}, \qquad a \in A, i \in \{1, \ldots, |A|\}$$

$$w_v \in \mathbb{R}, \qquad v \in V,$$

Numerical Experiments: Overview

- Long and short term experiments:
  – Different benchmarks
- Instances:
  – Shortest paths
  – Steiner trees
  – Also knapsack and abstract set cover.

Numerical Experiments: Benchmark

- Long term (Remember UCB1: $S_n \in \underset{S \in \mathcal{S}}{\operatorname{argmin}} \left\{ \bar{b}_{S,n} - \sqrt{2 \ln(n-1)/T_n(S)} \right\}$)
  - Extended UCB+

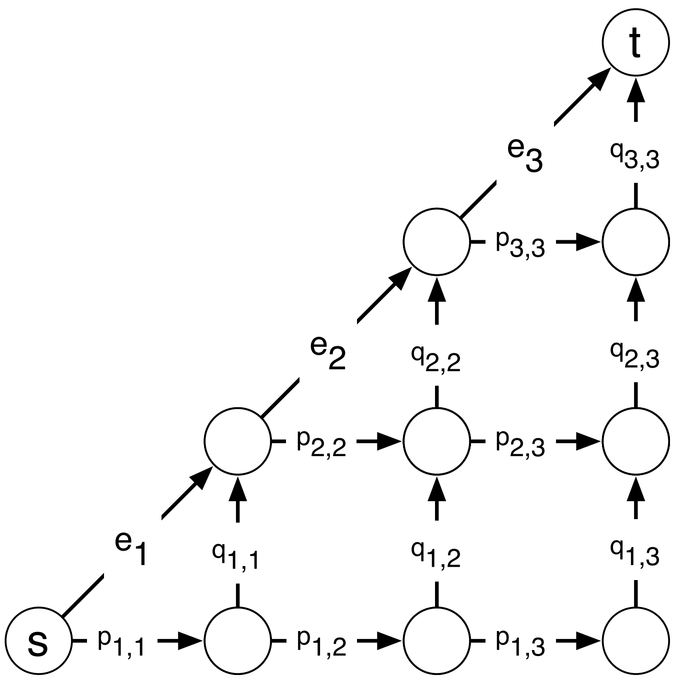
$$S_n \in \underset{S \in \mathcal{S}}{\operatorname{argmin}} \left\{ \max \left\{ \sum_{a \in S} \bar{b}_{a,n} - \sqrt{2 \ln(n-1)/(\min_{a \in S} \{T_n(a)\})}, \sum_{a \in S} l_a \right\} \right\}$$

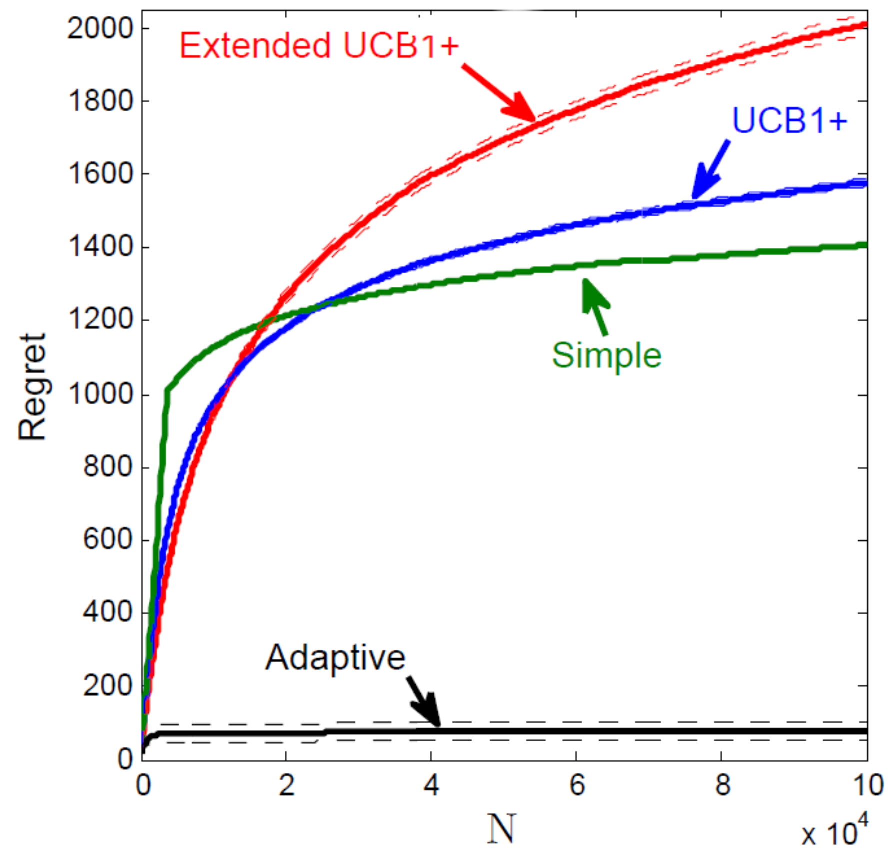
  - UCB+

$$S_n \in \underset{S \in \mathcal{S}}{\operatorname{argmin}} \left\{ \sum_{a \in S} \max \left\{ \bar{b}_{a,n} - \sqrt{(L+1) \ln(n-1)/T_n(a)}, l_a \right\} \right\}$$

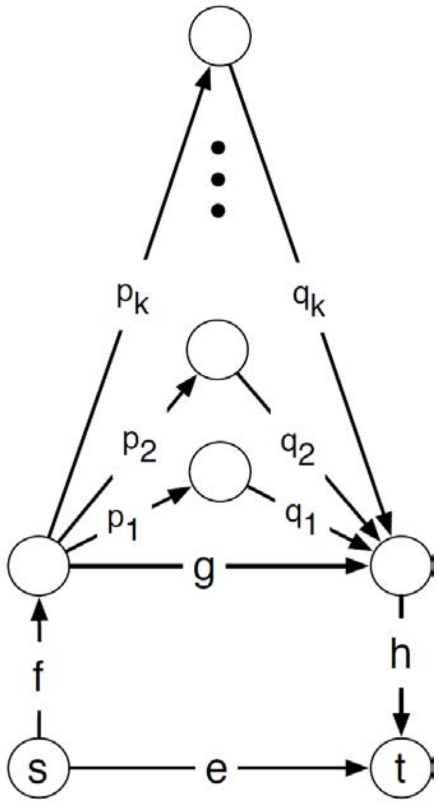
- Short term
  - Knowledge gradient (exponential-gamma) Ryzhov et al. (2012)
  - Gittins (Normal/Normal-Gamma) Lai (1987)

Long Term Experiments: Path 1

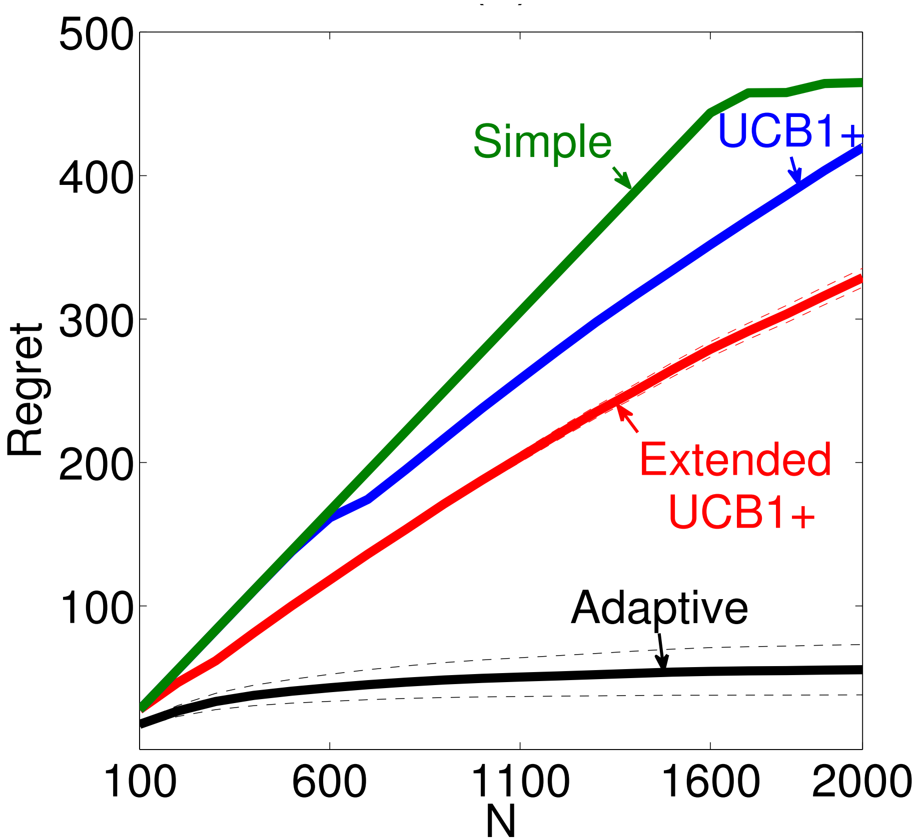




Long Term Experiments: Path 2

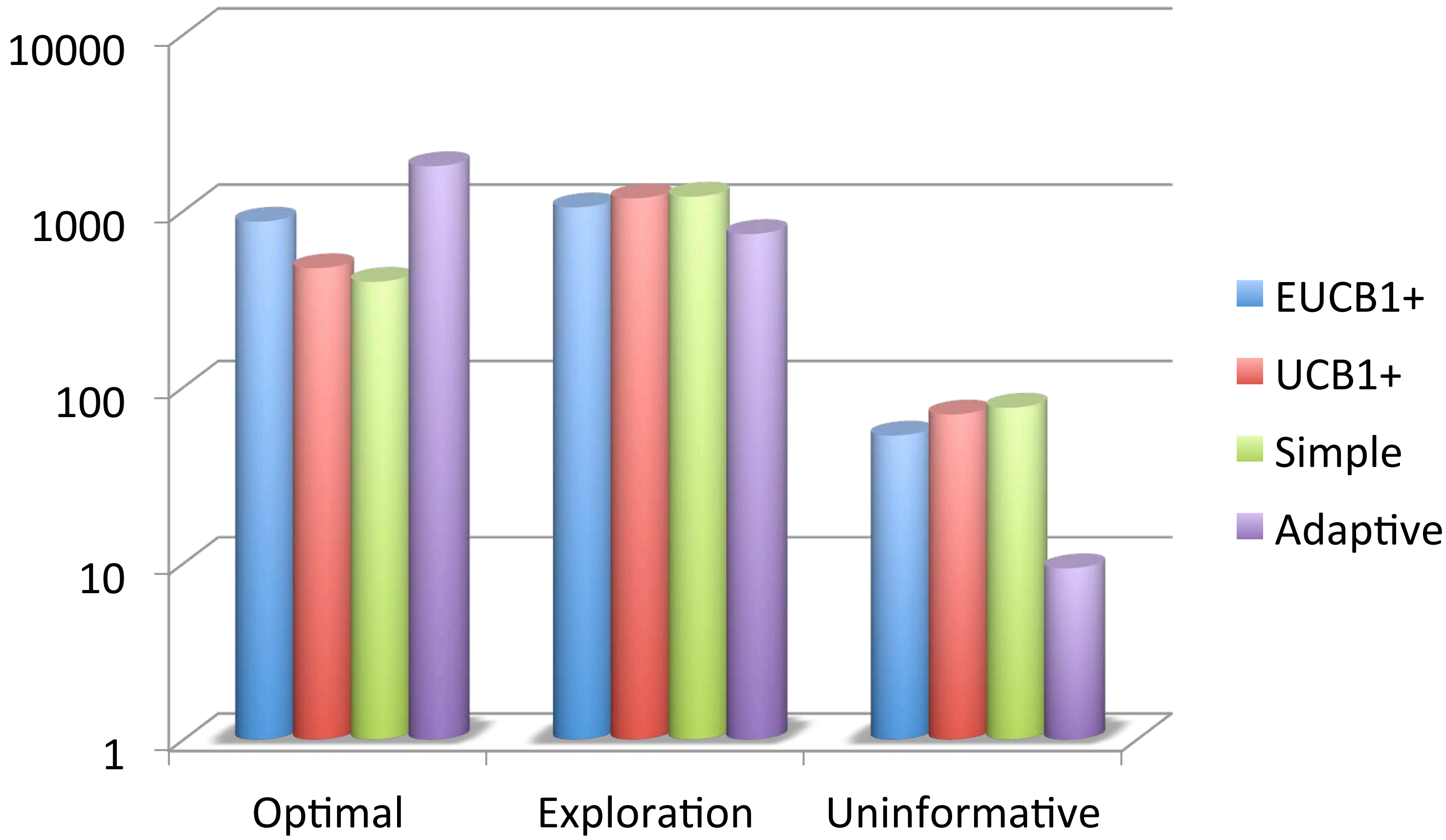


k=20



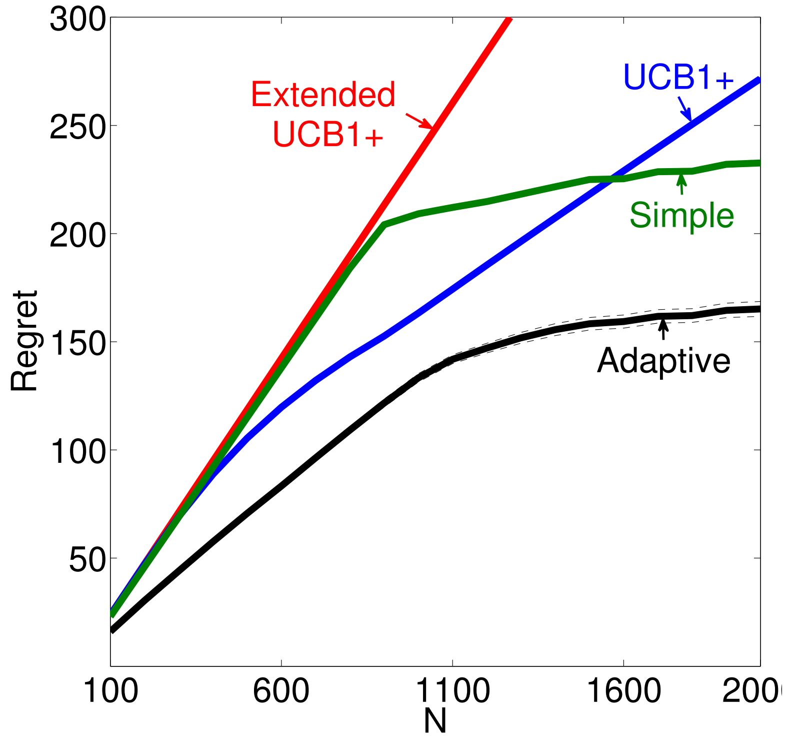
Long Term Experiments: Path 2

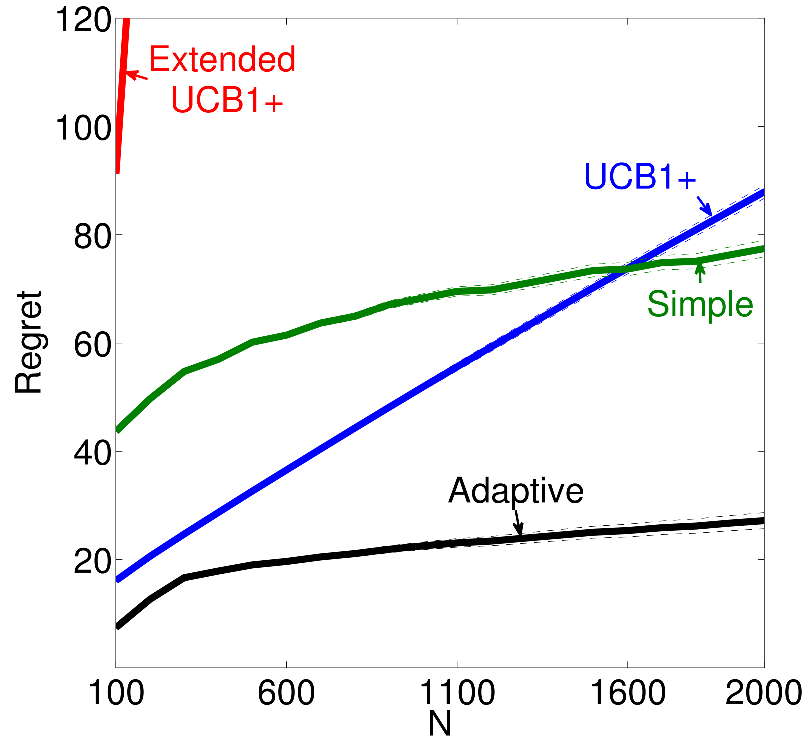
- Average number of selections for different arc classes.



Long Term Experiments: Path and Trees

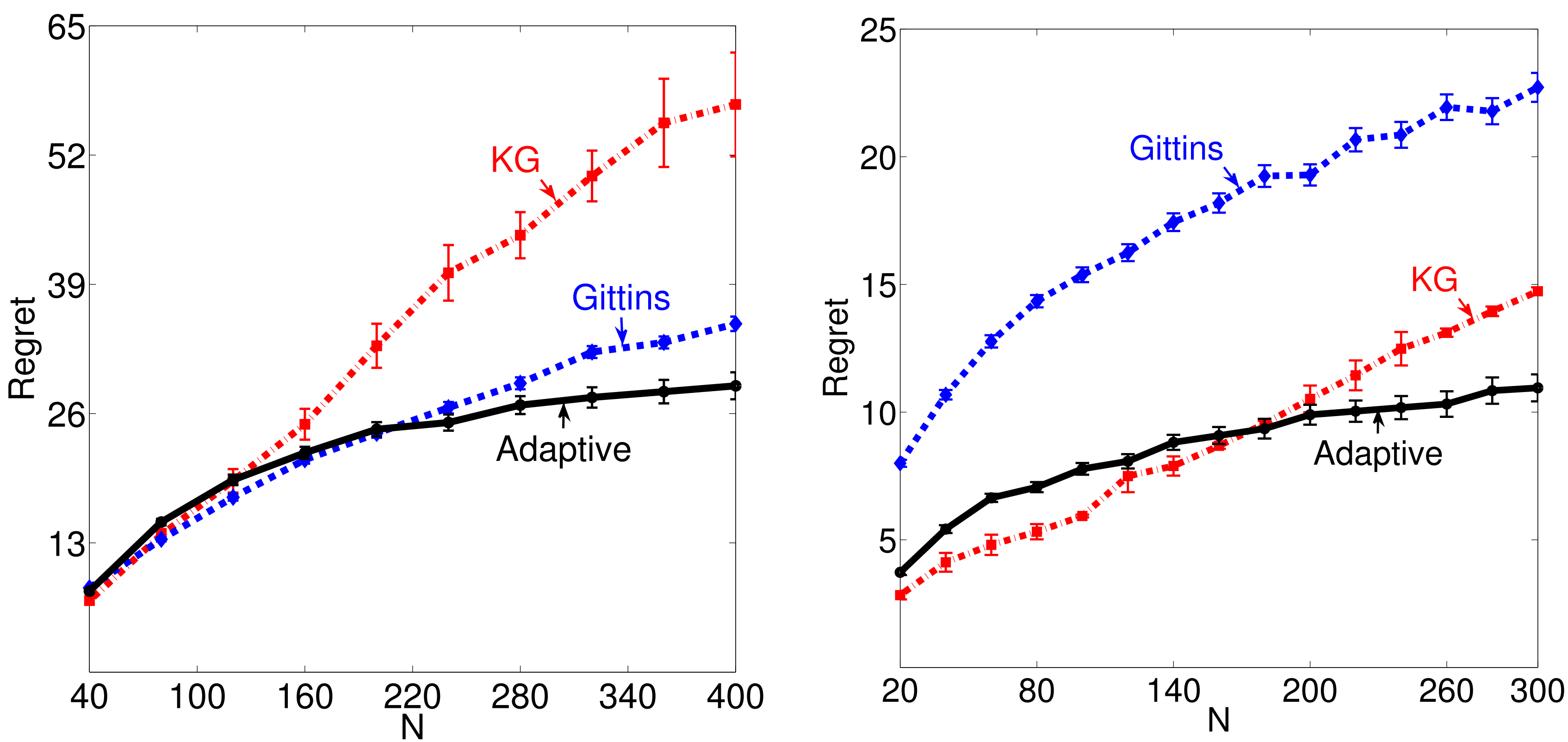
- Random Layer(5,4,3) graph (Ryzhov and Powell 2010)
- Steiner tree (|A|=18)





Long Term Experiments: Paths and Trees

- Random Layer(5,4,2) graph (Ryzhov and Powell 2010)
- Steiner tree (|A|=18)



Summary

- Traditional Exploration v/s Exploitation
  - **What** to exploit
  - **When** to explore
- Combinatorial Exploration v/s Exploitation
  - **What** to explore: critical elements
  - **How** to explore: optimality cover
- Implementable algorithm
  - Exploration/Exploitation cycles
  - Near-optimal long term performance
  - Competitive short term performance
- Complexity of OCP: new challenges

Limit on Achievable Performance

- Let $\mathcal{D}$ contain all subsets D of suboptimal elements that become part of every optimal solution if their costs are the lowest possible

- For any consistent policy π and set $D \in \mathcal{D}$

times element a tried

$$\lim_{N\to\infty} \mathbb{P}_F \left\{ \frac{\max\{T_{N+1}(a) : a \in D\}}{\ln N} \geq K_D \right\} = 1$$

distance between F and F'

- **What** needs to be explored? ***Critical subsets***

$$\mathcal{C} := \{C \subseteq A \,:\, \forall D \in \mathcal{D}, \exists\, a \in C \text{ s.t. } a \in D\}$$

Limit on Achievable Performance

- For any consistent policy π

$$\boxed{\liminf_{N\to\infty} \frac{R^\pi(F,N)}{\ln N} \geq \kappa(F)}$$

where

$$
\begin{array}{llll}
LBP: \quad \kappa(F) := \min & \sum_{S\in\mathcal{S}} \Delta_S^F \, y_S & & \textit{(min regret)} \\
\quad s.t. & \max\{x_a : a\in D\} \geq K_D, & D\in\mathcal{D} & \textit{(exp on critical subset)} \\
 & x_a \leq \sum_{S\in\mathcal{S}: a\in S} y_S, & a\in A & \textit{(solution cover)} \\
 & x_a, y_S \in \mathbb{R}_+, & a\in A, S\in\mathcal{S} & \textit{(non-negativity)}
\end{array}
$$

Proposed Policy

- For H_a such that for all $H > H_a$ and any $N > 0$

$$\boxed{\frac{R^{\pi_a}(F,N)}{\ln N} \leq G\ \Delta^F_{max} H + o(1)}$$

Size of minimal solution to OCP

- Gap in performance between lower and upper bounds

$$\kappa(F) \leq \frac{R^{\pi_a}(F,N)}{\ln N} \leq G\ \Delta^F_{max} H + o(1)$$