From Bandits to Experts: On the Value of Side-Observations

Shie Mannor
The Technion

Ohad Shamir*
Microsoft Research New England

NIPS

December 2011

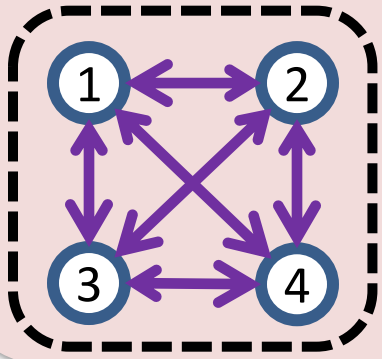
What We Do

- Sequential decision making: repeatedly choose among k actions
- **"Experts" Setting:** Learner sees rewards of all actions
- **"Bandits" Setting:** Learner sees only its own reward

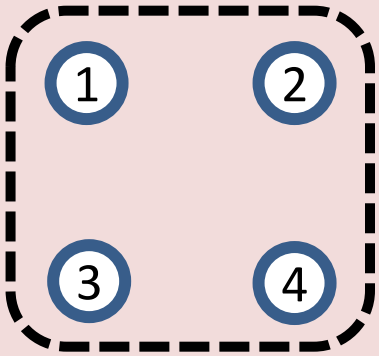
Our (More General) Model

By picking an action, Learner gets **side-information** on **some subset** of other actions. Captures:

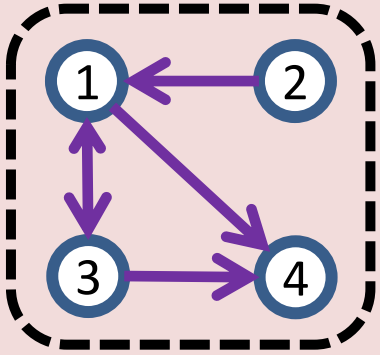
Experts…



…bandits…



…and in between!



Why

- Captures **structure** between the actions
  - Web Advertising



  - Sensor and communication networks

- **Related work:** Focus on affinity (Bandits in Metric Spaces) or stochastic correlations, while we make no such assumptions

Results

- New Efficient Algorithms
- Regret upper and lower bounds
  - Non-trivial dependence on information feedback structure
- Experiments
- Many open questions!

Come see our poster!