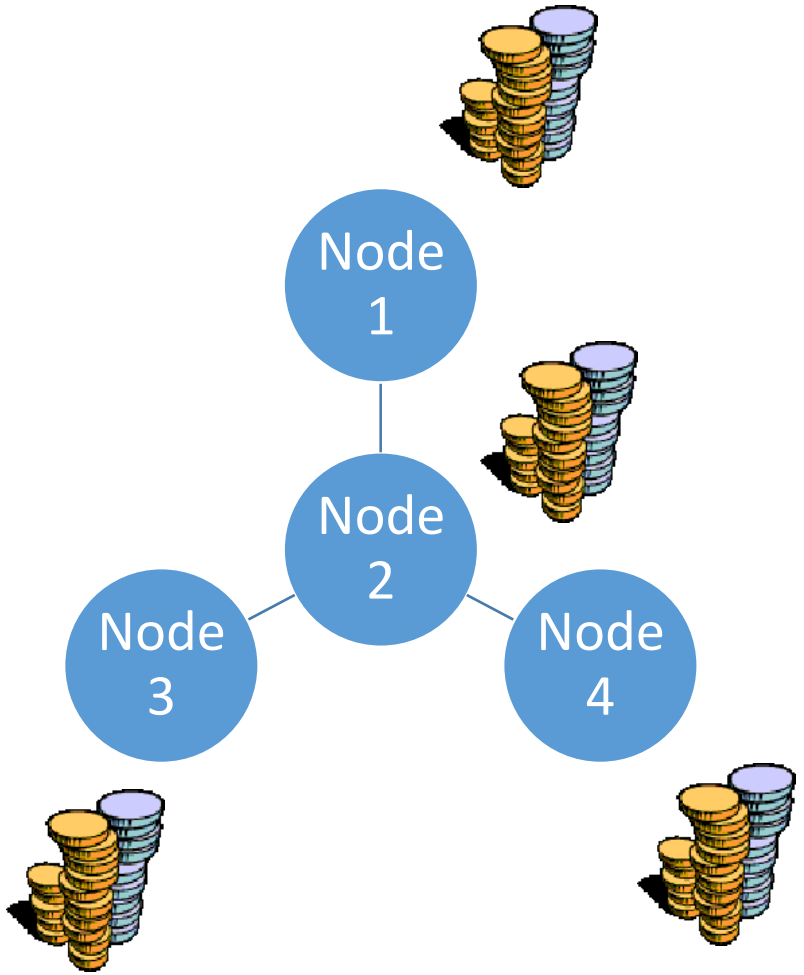
Distributed Multi-armed bandits

Milind Rao
CME323

Multi-armed bandits

- K arms or coins with different biases.
- At each time instant – toss any coin and observe reward. Explore-exploit
- Goal: Minimize regret of rewards
- Single player: UCB, Thompson Sampling, Online gradient descent
- Extend to: N nodes on a network, each playing and collaborating



A distributed MAB algorithm: Pregel framework

- w – probability distribution
- Each node picks a coin with distribution w, observes reward
- Node constructs unbiased estimate of reward g.
- Running estimate of cost vector – message
- A node mixes messages from neighbours, re-estimates distribution

Results

- Choosing step-sizes for different configurations
- Large message size