Randomly Sampling Maximal Itemsets
Sandy Moens and Bart Goethals
Frequent Itemset Mining

- Finding interesting patterns by e.g. support

- Problems:
  - Much redundancy
  - Many, many patterns
Frequent Itemset Mining

• Finding interesting patterns by support

• Problems:
  - Much redundancy
  - Many, many patterns
Frequent Itemset Mining

- Finding interesting patterns by support
- Problems:
  - Much redundancy
  - Many, many patterns
Pattern Set Mining

- Less redundancy
- Less patterns
- But: large enumeration space!
Pattern Set Mining

- Less redundancy
- Less patterns
- But: large enumeration space!

Step 1: Enumerate
Pattern Set Mining

- Less redundancy
- Less patterns
- But: large enumeration space!

Step 1: Enumerate  Step 2: Filter
Output Space Sampling

• No explicit enumeration
Output Space Sampling

- No explicit enumeration
Random Maximal Itemset Sampling

- Long patterns with low support
  - E.g. microarray data, recommendation

- Simple random walk over extensions
  - Quality measure $q$
  - Approximation measure $p$
Random Walk

Current Set $R$

A B C D E F G H I

Extensions $c_i \in C$

Rank $p(R \cup c_i)$
Random Walk

1. Sample

2. Prune by q

3. Rank by p
Random Walk
Spreading the Search

• Uniform Metropolis-Hastings
  - E.g. Hasan and Zaki, *Musk: Uniform sampling of k-maximal patterns* (SDM’09)

• Weight approximation score
  - Additive
  - Multiplicative
  - Adaptive
DEMO TIME