

Dynamic Interaction Graphs with Probabilistic Edge Decay

Session 25: Graph-based Data Management

Wenlei Xie (Cornell University)

Yuanyuan Tian (IBM Almaden Research Center)

Yannis Sismanis (Google, Inc.)

Andrey Balmin (Platfora, Inc.)

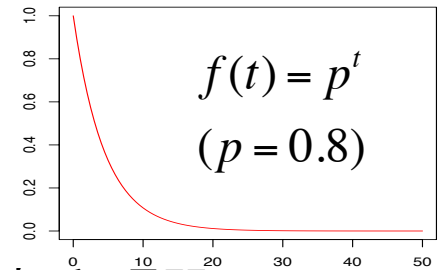
Peter J. Haas (IBM Almaden Research Center)

担当: 若林 啓 (筑波大学)

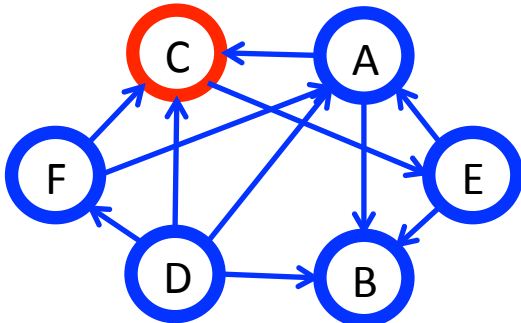
動的グラフの解析

– 時間とともにエッジの追加されていくグラフ

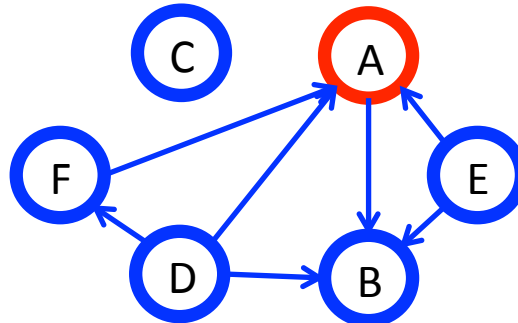
- 「新しい」エッジを優遇して解析したい (recency)
 - Sliding-windowは窓幅の大きさにセンシティブ
- 「新しさ」を減衰関数により重み付け (continuity)
 - 過去の全てのエッジがメモリ上に必要
 - 重み付けに対応できない手法もある
 - » クリーク発見, 連結要素抽出など



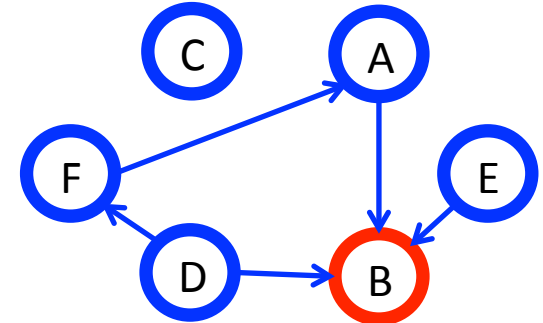
過去の全てのメンション



直近1ヶ月のメンション



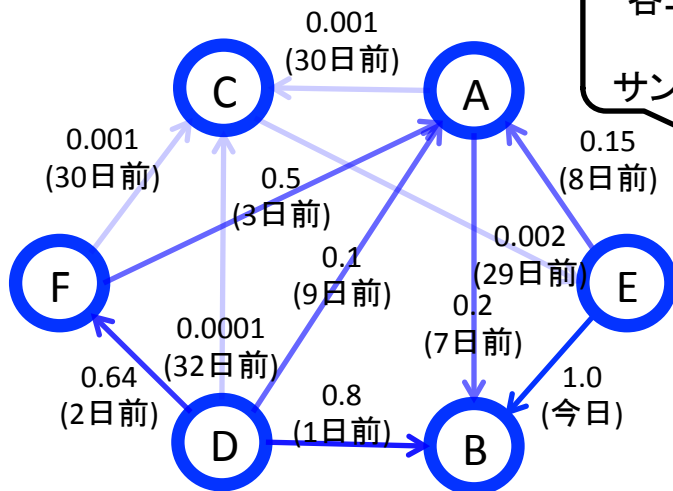
直近1週間のメンション



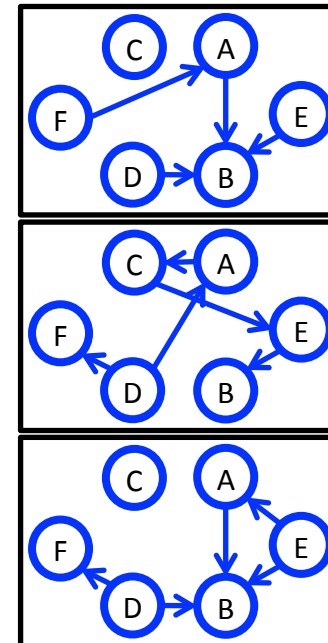
Uncertain graphとしての定式化

– 重みをエッジの存在確率と解釈

- サンプルングによりN個の重みなしグラフを得る
 - それぞれのグラフで解析アルゴリズムを実行し, 集約
- サンプルグラフの効率的な管理手法を提案
 - 更新の度に元のUncertain graphを計算していたのでは非効率



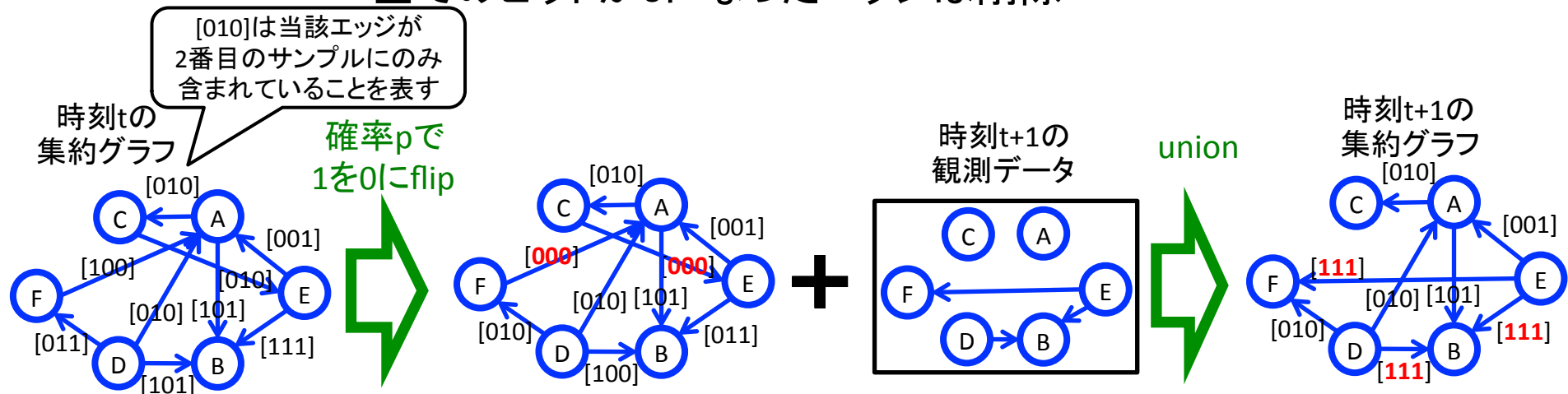
各エッジの確率は
独立と仮定
サンプリングは容易



集約グラフと差分更新

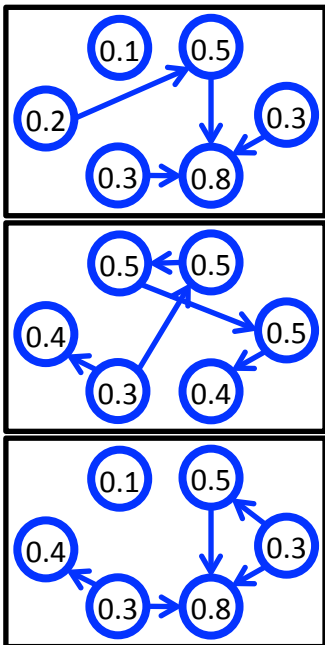
- N個のサンプルを素直に保持するのは無駄
 - エッジごとにNビット配列を持つ単一の集約グラフで保持
- 元の確率グラフを忘れてサンプルを差分更新
 - 各時刻で1が立っているビットを確率pで0にフリップ
 - 元の分布からの正しい(unbiasedな)サンプルになる

$$f(t+1) = p^{t+1} = p \times p^t$$
 - 全てのビットが0になったエッジは削除

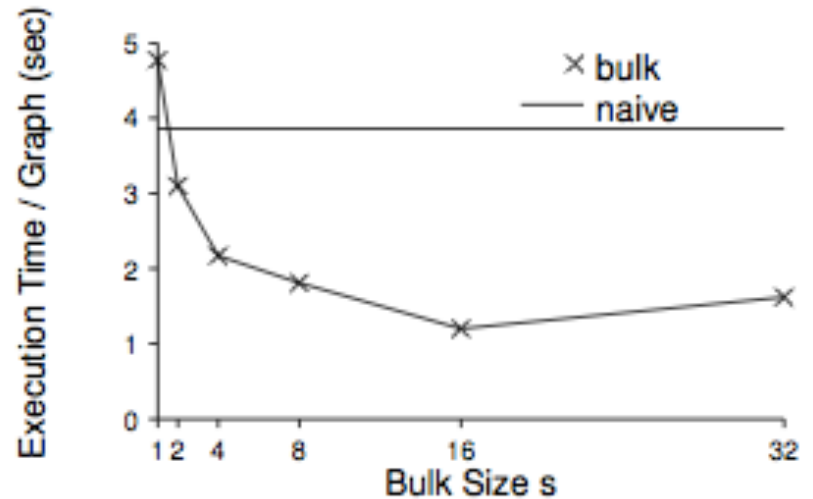
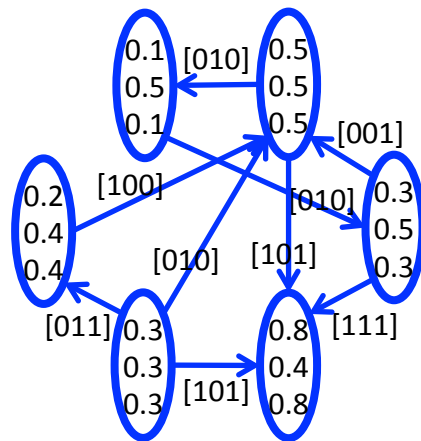


Bulk Graph Analysis

- グラフ解析アルゴリズムも集約グラフ上で行う
 - ジョブ開始のオーバヘッドやキャッシュ・圧縮効果で有利
- Apache Spark + GraphX plugin の上に実装
 - 16台のクラスタ, Twitterのメンショングラフで実験
 - サンプル1つずつ解析するよりも4倍の高速化



==



論文のFig. 9 (a)