

# **Convex Risk Minimization to Infer Networks from Probabilistic Diffusion Data at Multiple Scales**

Emre Sefer and Carl kingsford

# **False Rumors Detection on Sina Weibo by Propagation Structures**

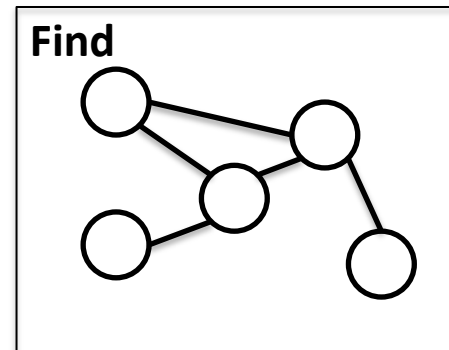
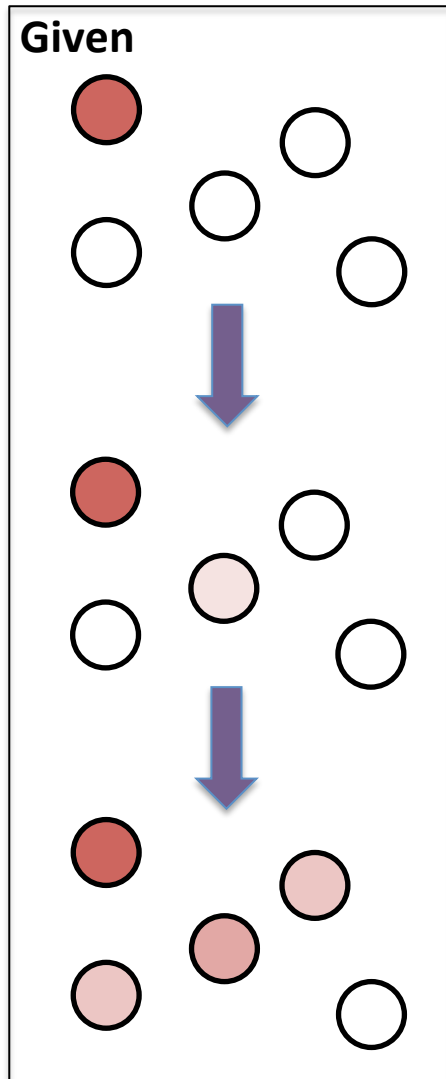
Ke Wu, Song Yang, and Kenny Q. Zhu

# **Identifying Users in Social Networks with Limited Information**

Knorases Vesdapung and Hector Garcia-Monica

# **CONVEX RISK MINIMIZATION TO INFER NETWORKS FROM PROBABILISTIC DIFFUSION DATA AT MULTIPLE SCALES**

# Problem – Link Prediction(?)



**Given:** 感染の不確実時系列情報

**Find:** エッジ集合

# Background

## 問題の重要性:

- 感染の観測は容易だがエッジの観測は困難な場合が多い
- 感染の観測は通常不確実
  - 熱と咳があるからといって必ずインフルエンザというわけではない

## **SEIR Model:** Susceptible-Exposed-Infected-Recovered Model

- 本研究で扱う感染モデル
- SI, SIS, SIR model を特殊ケースとして含む
- Independent cascade

# Method

**CORMIN**: Convex Risk Minimization to Infer Networks

- 凸最適化問題として解く
- 観測データは不確実性を持つためR1正則化項を持つ

## **Micro-scale inference:**

(例) 人の感染情報から接触ネットワークを推定

## **Macro-scale inference:**

(例) 州の感染情報から人の移動ネットワークを推定(?)

# Contributions

## **Probabilistic modeling**

- 感染状態を確率分布で表現:  $P(s)+P(e)+P(i)+P(r)=1$
- **R1**正則化

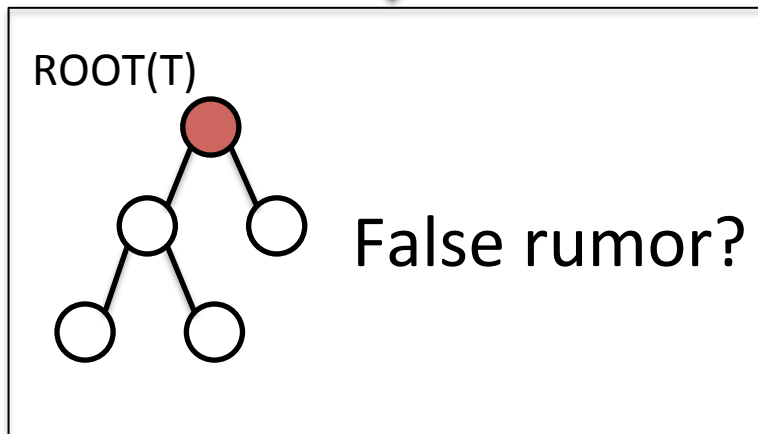
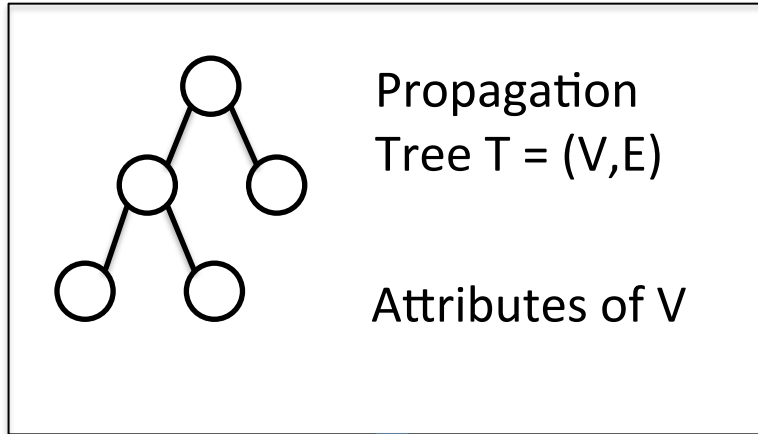
## **Modeling both edge existence and non-existence**

## **Inferring both micro and macro scales**

- 人から人への感染経路
- 州から州への感染経路

# **FALSE RUMORS DETECTION ON SINA WEIBO BY PROPAGATION STRUCTURES**

# Problem – Tree classification



**Given:** Tree  $T = (V,E)$

**Find:** Whether or not  $ROOT(T)$  is a false rumor



# Background

## **問題の重要性:**

- 明らか？

## **Rumor:**

- 広く伝搬しているが正しいことが確認されていない情報

## **False Rumor:**

- Rumorのうち最終的に正しくないことが確認したもの

# Method

いろいろ特徴抽出してSVMにぶち込む

## 工夫した点:

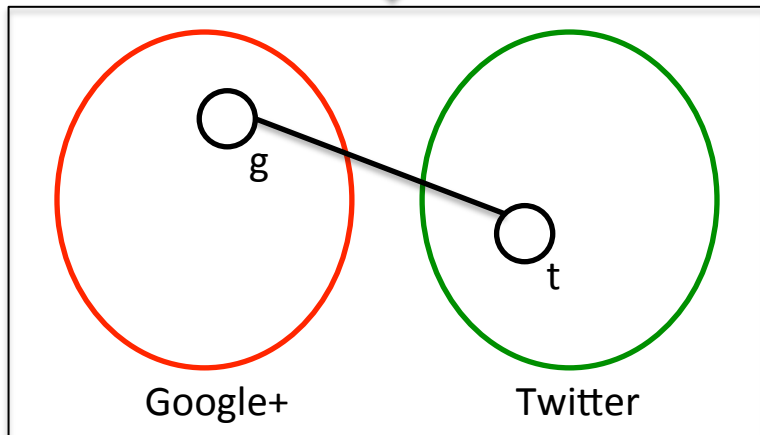
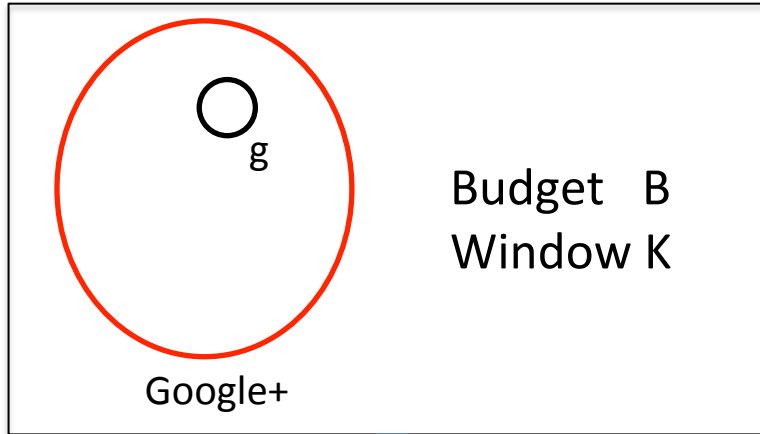
- テキスト情報だけでは分類が難しいため、グラフカーネル (Random walk kernel) を使う

# Contributions

- Message propagation を tree でモデル化(?)
  - なにも新しく見えない
- Random walk graph kernelを提案(?)
  - 既存のものを、離散でなく連続のラベルも使えるように拡張しただけ
- いろんな特徴量をSVMに入れたら精度が上がった
  - まあ、はい。

# **IDENTIFYING USERS IN SOCIAL NETWORKS WITH LIMITED INFORMATION**

# Problem – Entity Resolution



**Given:** Node  $g$   
Budget  $B$   
Window  $K$

**Find:** 期間  $K$  中に費用  $B$  のAPI call  
を使って  $g$  と同じユーザ  $t$  を見つ  
ける

# Background

## 問題の重要性:

- Twitterをはじめとして多くはAPI制限がある
  - → Limited Information
- 様々な応用
  - $g$  と  $t$  に重複して余計な広告をうたなくて済む
  - 複数のSNSからのテロリストの情報収集
  - 同じネットワークからの同一ユーザの検出
  - その人 ( $g, t$ ) についてより多くの情報を収集

# Method

**1. Get the potential candidates for  $g$ .**

**2. Exploration:**

$g$  の候補集合を評価するために最良と思われる情報を費用  $B$  の許す限り集める

**3. Evaluation:**

集めた情報を用いて候補集合を評価し、最も  $g$  と同一であると思われるノードを  $t$  とする

# Contributions

- **問題定義**
  - social network ER with limited resources
- **既存手法を拡張して定義した問題に適用**
- **TwitterAPIについて、どの情報を得たらより効果的にERできるか分析**
- **マッチングアルゴリズムを二つ提案**
- **実データで評価**