Algoridam

Théorie algorithmique de nouveaux modèles de données

IRIF
Claire Mathieu
coordinatrice

Sophie Laplante
Frédéric Magniez
Adi Rosén
Michel de Rougemont
Miklos Santha
Olivier Serre
total 128 mois

12 mois de postdoc

DIENS
Chien-Chung Huang
partenaire local

Pierre Aboulker
Tatiana Starikovskaya
total 48 mois

12 mois de postdoc

LIP6
Christoph Dürr
partenaire local

Spyros Angelopoulos
Evripidis Bampis
Vincent Cohen-Addad
Bruno Escoffier
total 84 mois

12 mois de postdoc

Overview

The traditional computation model:

input —> algorithm —> output

no longer fits most applications

New challenges: massive, scattered, disorganized, erroneous, constantly evolving data input.

Theory has fallen behind practice

Need theoretical foundations for the types of algorithms that have emerged

our approach: axiomatic perspective, theoretical models, rigorous theorems

Focus on three settings

A Massive data (data that is too large to fit into memory)

B Noisy data (e.g. the data cannot be reliably accessed and is hence observed with error or noise)

C Dynamic data (e.g. the data evolves constantly)

Snapshots: three examples

Workpackage A: Massive Data
Task 2: Streaming algorithms
Open problem: approximate pattern matching

find all substrings of a given longer string (text) that are close to a given set of short strings (patterns)

motivation : text analysis and bioinformatics

extensively studied for Hamming distance when there is a single pattern

goal: develop streaming algorithms for multiple patterns & other distances

Workpackage B: Noisy and Unreliable Data
Task 5: Data with intervals of uncertainty
Open problem: linear programs with uncertain constraints

input: linear program

min cx subject to $Ax \geq b$ and $x \geq 0$

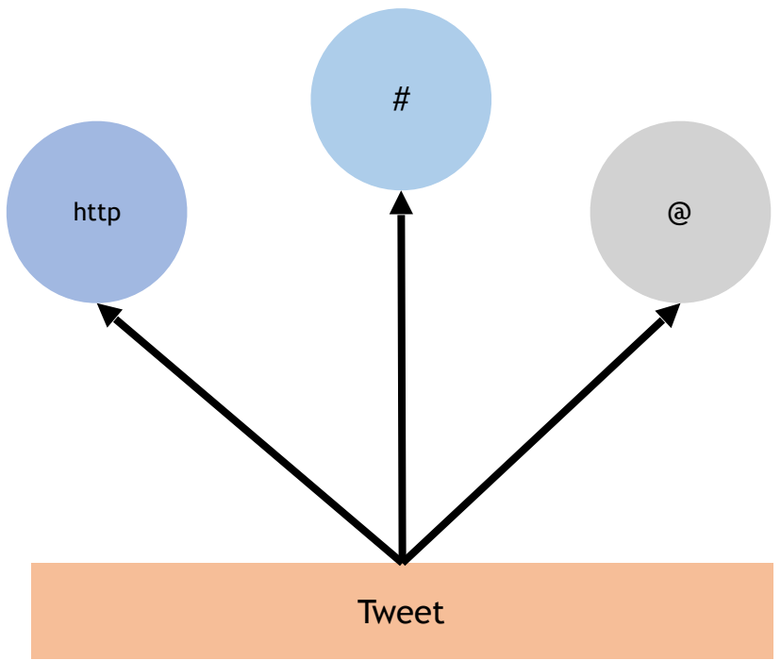
s.t. right hand sides are uncertainty intervals

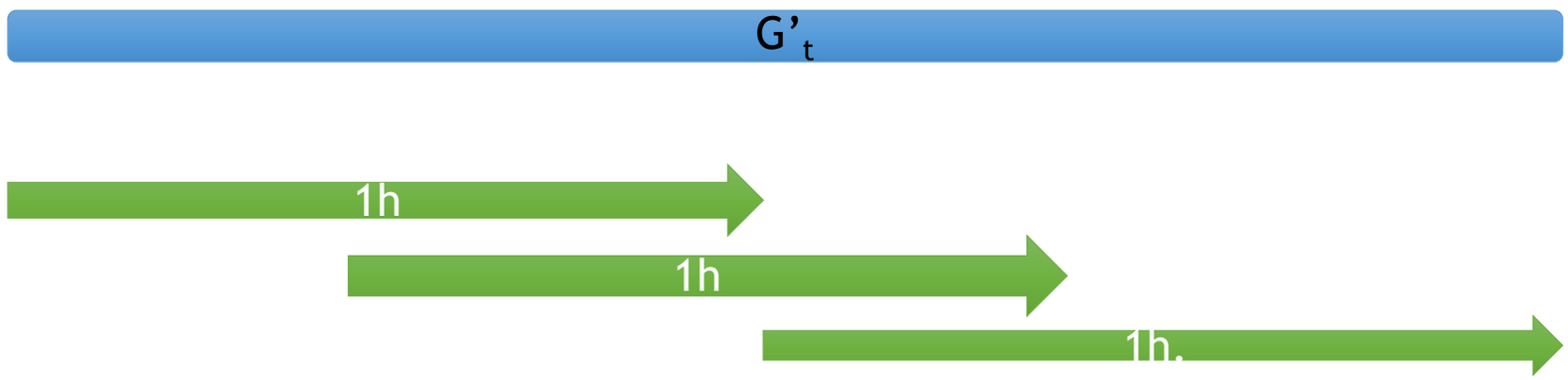
output: optimal solution x

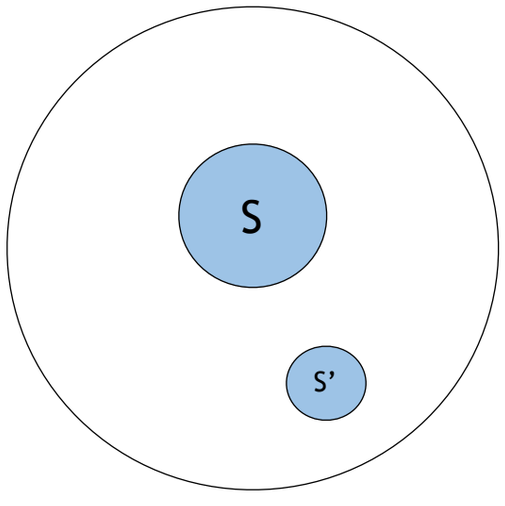
goal: query as few right hand sides as possible

models: problems where the exact resource capacities are only known after querying —> costly inventory or accounting procedures

Workpackage C: Dynamic Data
Task 9: Evolving graphs
Open: Detecting clusters in evolving Twitter streams







Let R(k) a Reservoir with k edges

Detection Algorithm. Let C the largest connected **component of R(k). If > 10, then Accept else Reject.**