BERLIN INSTITUTE OF TECHNOLOGY, MACHINE LEARNING DEPARTMENT

# The Title of Your Fab Research Project

**Theoretical Lab Rotation Report**

Ivana Kajić
BCCN Berlin

Supervisors:

June 14, 2014

**Abstract**

The dynamics of published and shared content in social networks reveals much about the underlying network structure. One important aspect of such networks are users who publish content before the majority of community starts to talk about it and share it. We regard them as high impact users because of their influence in terms of spawning retweets and mentions.In this lab rotation, we use machine learning methods to detect high impact users in the social network Twitter. To this end, we implemented an online, content-based learning approach known as canonical correlation analysis. CCA uses gradient descent optimization to update its parameters and is therefore capable of dealing with large amounts of data in realtime.

## 1 INTRODUCTION

### 1.1 THE TASK

## 2 RELATED WORK

## 3 METHODS

DATA AND FEATURES

$$X_u = [x_u(t=1), \ldots, x_u(t=T)] \in \mathbb{R}^{W \times T} \tag{3.1}$$

**Algorithm 1** Canonical Trend Algorithm

**Input:** Users $\mathscr{X}_u,\ u = 1,\dots,U$, learning rate $\eta_0$
# For each new sample $x_u(t)$
**for** $t = 1$ **to** $T$ **do**
  # Loop over all users
  **for** $u = 1$ **to** $U$ **do**
    $y_u(t) = 1/(1-U)\sum_{u'\neq u} x_{u'}(t)$
    # Temporal Embedding
    $\tilde{x}_u(t) = [x_u(t-N_\tau)^\top,\dots,x_u(t-1)^\top]^\top$
    # Stochastic gradient update of canonical directions $w_{\tilde{X}},\ w_y$
    $w_{\tilde{X}} \leftarrow w_{\tilde{X}} + \eta_0/t\ \ \tilde{x}_u(t)y(t)_u^\top w_y$
    $w_y \leftarrow w_y + \eta_0/t\ \ w_{\tilde{X}}^\top \tilde{x}_u(t)y(t)_u^\top$
    # Normalize canonical directions
    $w_{\tilde{X}} \leftarrow w_{\tilde{X}}\|w_{\tilde{X}}\|_2^{-1}$
    $w_y \leftarrow w_y\|w_y\|_2^{-1}$
  **end for**
  Rank Users $\propto$ Corr$(w_{\tilde{X}}^\top x_u, w_y^\top y_u)$
**end for**

$$Y_u = 1/(U-1)\sum_{u'\neq u} X_{u'} \in \mathbb{R}^{W\times T} \tag{3.2}$$

### 3.1 Canonical Trend Analysis

## 4 Results