Thematic Patterns in Georeferenced Tweets through Space-Time Visual Analytics

Piotr Jankowski

Department of Geography

San Diego State University

Why analyze microblogs such as Tweets?

- Validation of socioeconomic theories

- Location-specific marketing

- Timely and place-specific information from "social sensors"

- Prediction of social phenomena

# Challenges in microblog analysis

- High number of user-generated tweets; worldwide more than 340 million/day
- High level of noise (i.e. unrelated messages, abbreviations, typos, and plain nonsense)

Objectives

- Discover dominant microblogging themes within a specific geographic area
- Find their spatial and thematic distribution patterns

Data

- 300k georeferenced tweets from greater Seattle area from August to October 2011
- 13.7k Twitter users
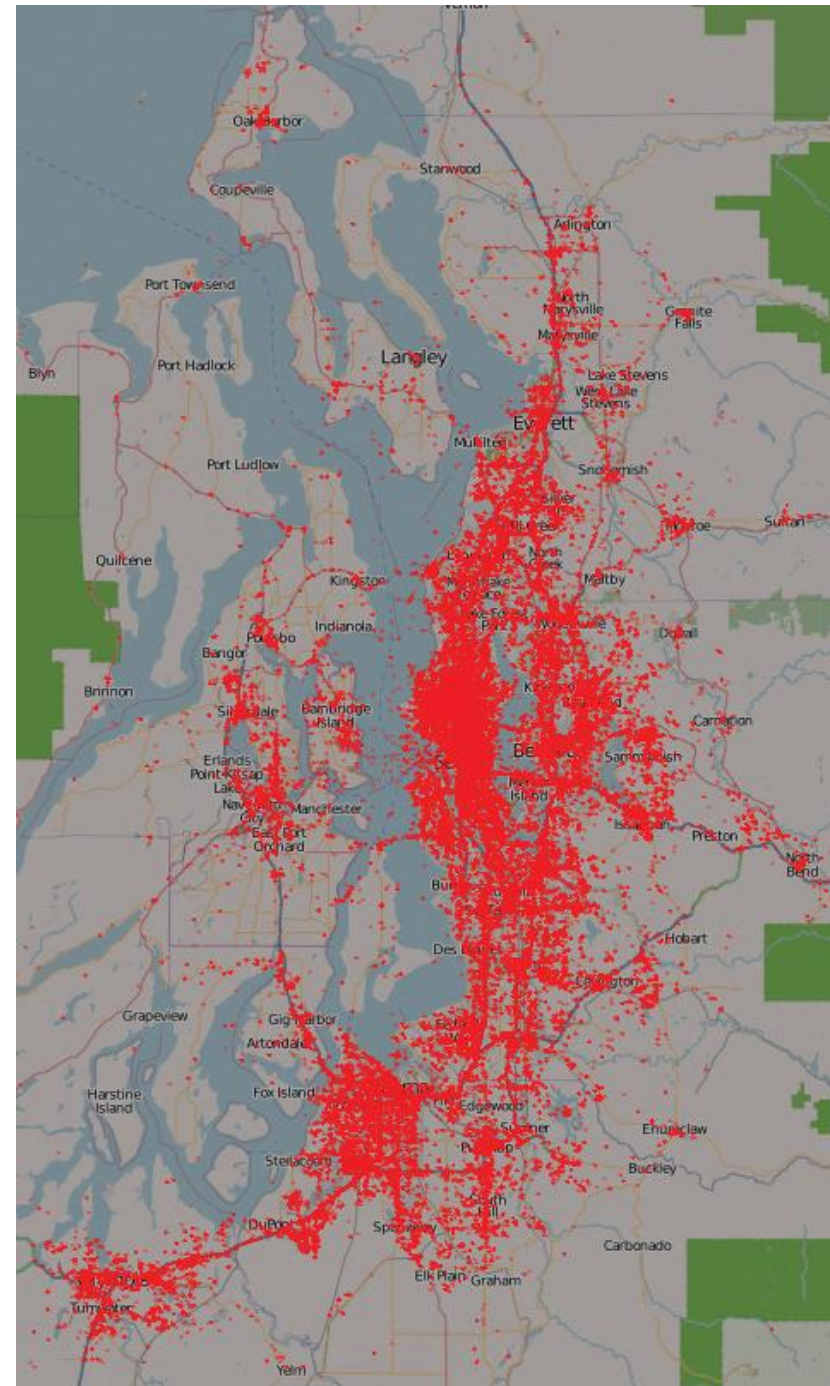
Focus on Tweeting Behavior of the Locals

- Concentrate on tweets of Seattle locals (exclude visitors)
  - Count for each unique user ID the days inside the greater Seattle area versus the days outside within the 60-day period
- 163K of tweets from 2.6k 'local' Twitter users

Content Analysis

- List of common topical categories
- 20% of analyzed tweets contain
  one or more topic-related keywords

Term	Frequency
Food	6,247
Love	4,074
Family	3,767
Work	3,076
Education	2,407
Home	1,954
Private event	1,928
Music	1,850
Sports	1,704
Game	1,678
Friends	1,410
Health	1,358
Coffee	1,136
Transport	1,120
Fitness	1,050
Alcohol	981
Weather	925
Sweets	876
Money	524
Public event	345
Tea	214
Wellness	151

Spatial Patterns of Tweets: a multi-topical distribution

- Density-based clustering of tweet locational origins (source locations)
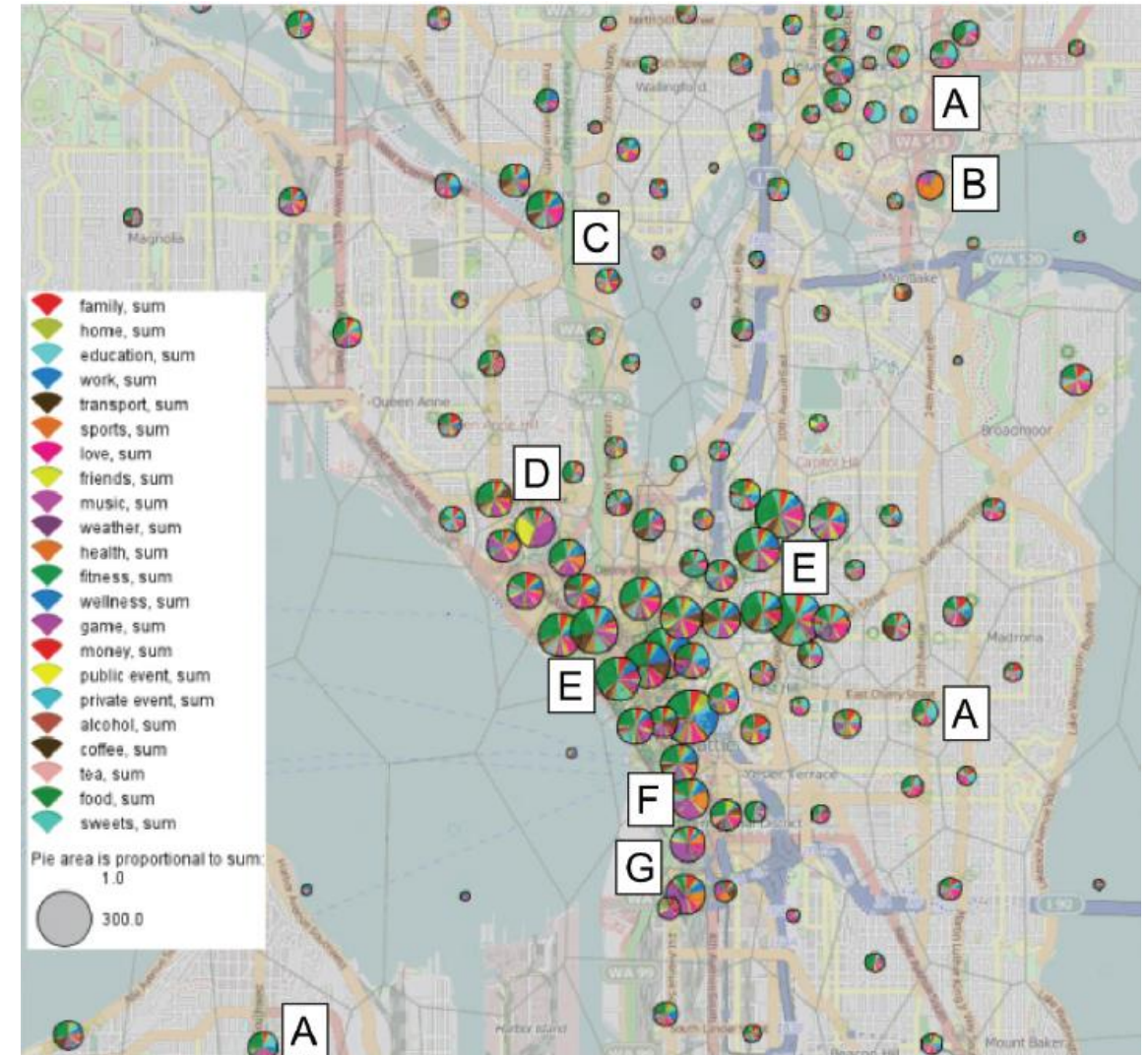
  A – education

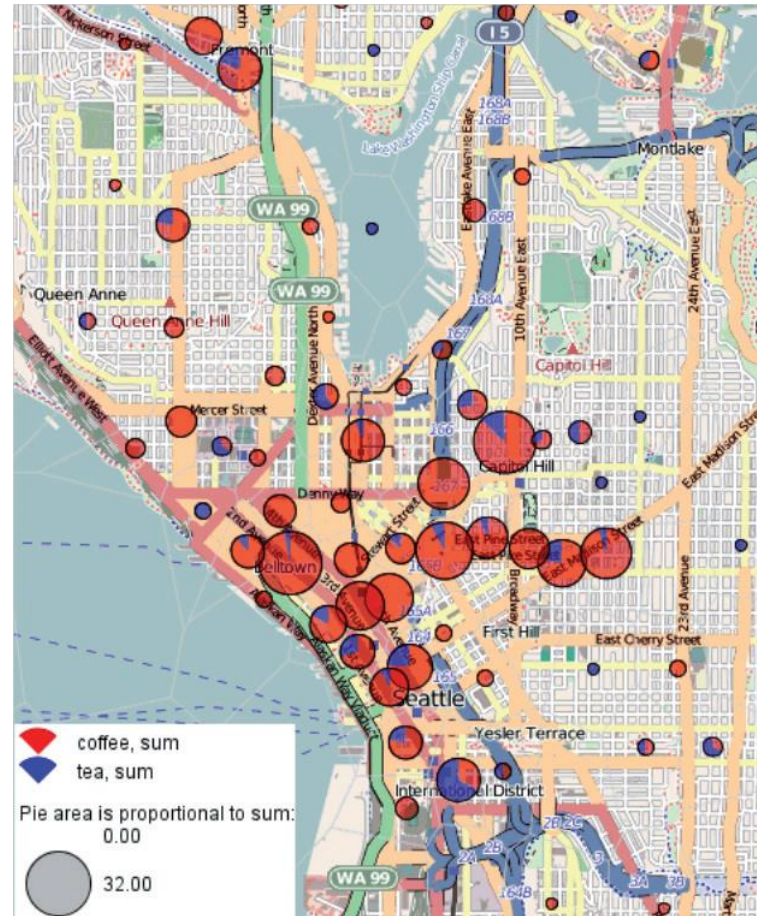 B – sports

  C – love

  D – music

  E – coffee
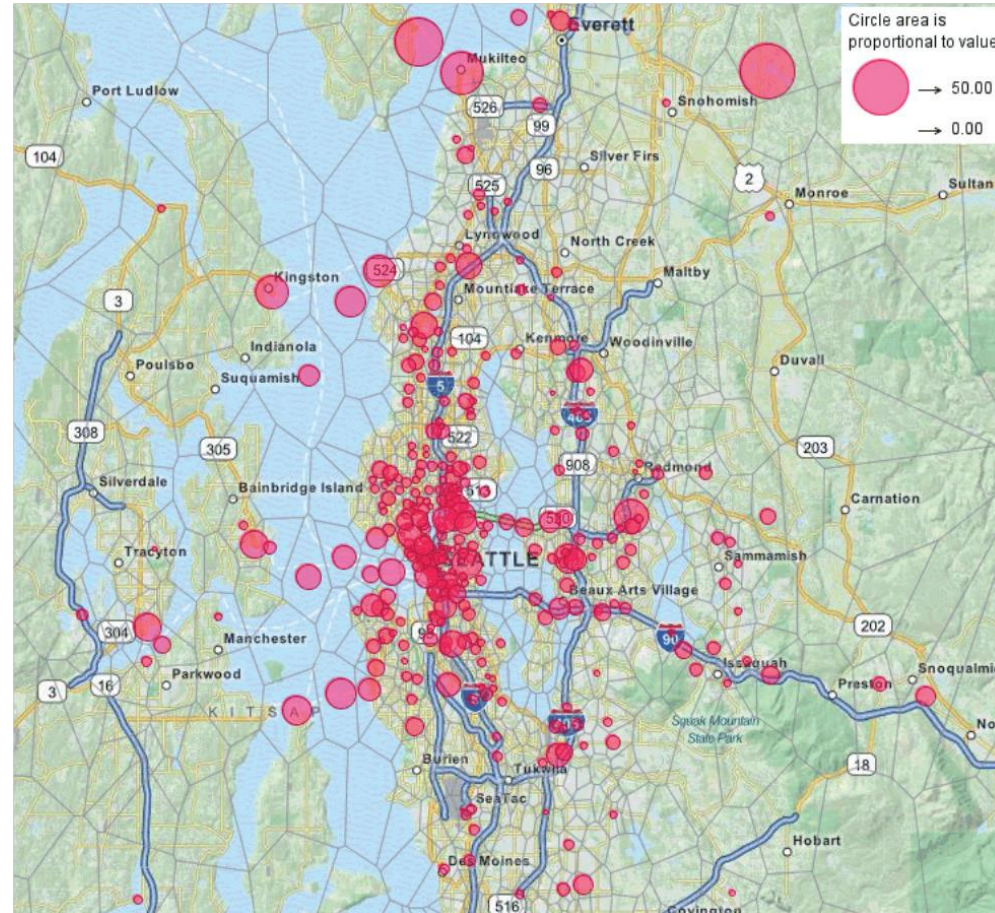
  F – sports & music

  G – sports & game
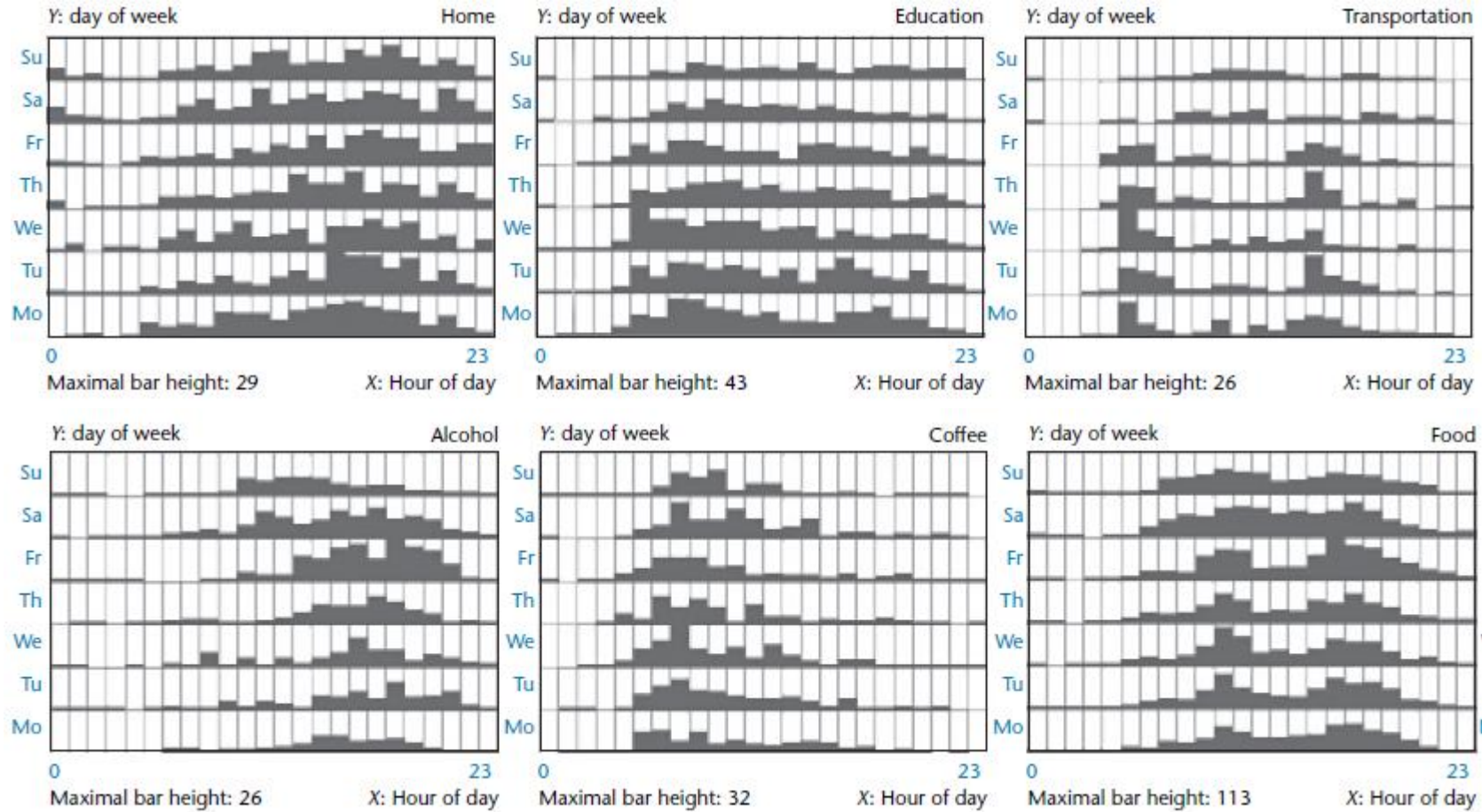
Spatial Patterns of Tweets: a bi-topical distribution

Spatial Patterns of Tweets: a single topic distribution - *transportation*

Temporal Patterns of All Tweets

Y: day of week

X: hour of day

Maximal bar height: 148 clusters

Temporal Patterns of Tweets on Selected Topics

Acknowledgement

THEMATIC PATTERNS IN GEOREFERENCED TWEETS THROUGH SPACE-TIME VISUAL ANALYTICS

By Gennady Andrienko, Natalia Andrienko, Harald Bosch, Thomas Ertl, Georg Fuchs, Piotr Jankowski, and Dennis Thom