

LA Grid Summit 2011 November 3rd & 4th 2011 **Florida Atlantic University**

FLORIDA INTERNATIONAL UNIVERSITY



Partnership for International Research and Education A Global Living Laboratory for Cyberinfrastructure Application Enablement

Describing Spatial Changes on Micro-blog Activity

Eduardo Ruiz, PhD Candidate, UCR **FIU Advisor:** Vagelis Hristidis, UCR

PIRE International Partner Advisor: Aris Gionis, Carlos Castillo Yahoo! Research Barcelona.

National Science Foundation

Latin American Grid

. Research Overview and Outcome

Overview

• Previous models study how the activity on twitter change on time.

• Idea: There are hash-tags that change their location from one day to other? Can we measure this change?





•Applications: Knowledge Mining. Relations in this domain can be different to the ones in the time model.

• Examples: Findings tags that are static, tags with a pattern of movement, flocks (tags that move together.

• Outlier movements have always an explanation?



recast is calling for an 80% chance of #snow this Wednesday the possibility of 12 inches in the central Rockies. Bring it!

Jake S610 Jake Silverman High tomorrow of 75. Tuesday to wednesday night, 8-16 inches of

J2Ski Snow J2Ski Snow Reports #Snow - At least 18 Ski Resorts near #Bolzan within 48 hours. - i2ski.com/z/hi

s_NeeSuh_Duh anissia ransom nyBody ready 4 the #snow ???? I am!!!

Activity Map for Tweets for #snow on February 4th

What	is a	movement?

• A movement is any change on the location of the activity that can be detected visually.

• Movements on this domains are not linear or cohesive. (Tracking hurricanes)

- Two task:
 - Detect the movement.

• Establish a measure for Distance • Select the top-k using this measure or use a threshold.

Summarize Movement

•Idea: Use the hierarchy to present the most important information. •Compression: find two sets of Algorithm 1 FindSummary(n,f,k) if n is a cell then for $u \in ancestor(n)$ do $ost_n^0(u) = |f(n) - avg(u)|$ $summ_{n}^{0}(u) = \{\}$ for $j \in 1 : k \operatorname{de}$ $ost_n^j(u) = 0$ $summ_n^j(u) = \{n\}$ end for end for $cost_n^0(\bot) = \infty$ $summ_n^0(\bot) = \emptyset$

Data Selection

• Select all the tweets between 01/01/11 and 03/01/11

• Filter those tweets that have at least one hash-tag.

• Localize each tweet using geo-location coordinates or heuristics.

• One entry for each hash-tag with the emission date, latitude and longitude.

Results

• Track two hurricanes (Earl – Bonnie) using tweets. Applied a linear model to predict trajectories.

• User evaluation. User is presented with animations and click which one they prefer. • EMD is the best ones as consider multiple centers and







regions that represent the initial state and the movement (video compression).

•Given a budget of k find in efficient way a summary. We propose a dynamic programming adaptation of the Facility Location.



for $l \in 0 : p$ do $FindSummary(child_p, k, k)$ for $j \in 0 : k$ do **r** $u \in ancestor(n) \cup \{ \perp \{ \mathbf{do} \\ (\delta_j(u), \sigma_j(u)) = BestSubTrees(u, j) \}$ $\sigma_{j-1}(n), \sigma_{j-1}(n) = BestSubTrees(u, j-1)$ $\delta_i(u) < \delta_{i-1}(n)$ then $st_n^j(u) = \delta_j(u)$ $umm_n^j(u) = \sigma_j(u)$ $cost_n^j(u) = \delta_{j-1}(n)$ $summ_n^j(u) = \sigma_{j-1}(n) \cup \{n\}$

more coherence.

• Single center are sensitive to

small changes

• Summarization results on work. • Some results on movement prediction.



Conclusions and Future Work

• We define the problem of movement detection for topics.

• We study some of the differences of this problem with other works on spatial data

• We propose a adaptation to summarize the movement using hierarchy.

•Future work: focus on the applications of the movement detection, in particular movement prediction and clustering.

II. International Experience Munich in two days Dachau, Munich Center,

Bavarian Palace.





Barcelona: not the same

- More focused on my work have less time to be around.
- Focus on missed locations and activities.
- Still need a another trip :)
- •Best trip: Monserrat



Yahoo Research Shorter experience (lot of work). •Why go there: Access to datasets and tools. Faster response time. • What is the current status: Collaboration on Skype (weekly meetings). Working on submission.

III. Acknowledgement

The material presented in this poster is based upon the work supported by the National Science Foundation under Grant No. OISE-0730065. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.