Detection and Delineation of Events and Sub-events in Social Networks

A. Saravanou*, I. Katakis+, G. Valkanas#, D. Gunopulos*

*National and Kapodistrian University of Athens, +University of Nicosia, #Detectica

ICDE PARIS 2018

MOTIVATION

In real life, a series of highlights (sub-events) constitute an event

E.g., a soccer game can be described by a sequence of sub-events (goals, penalties, cards, halftime, etc) or a music show can be described by the sequence of the artists' appearances

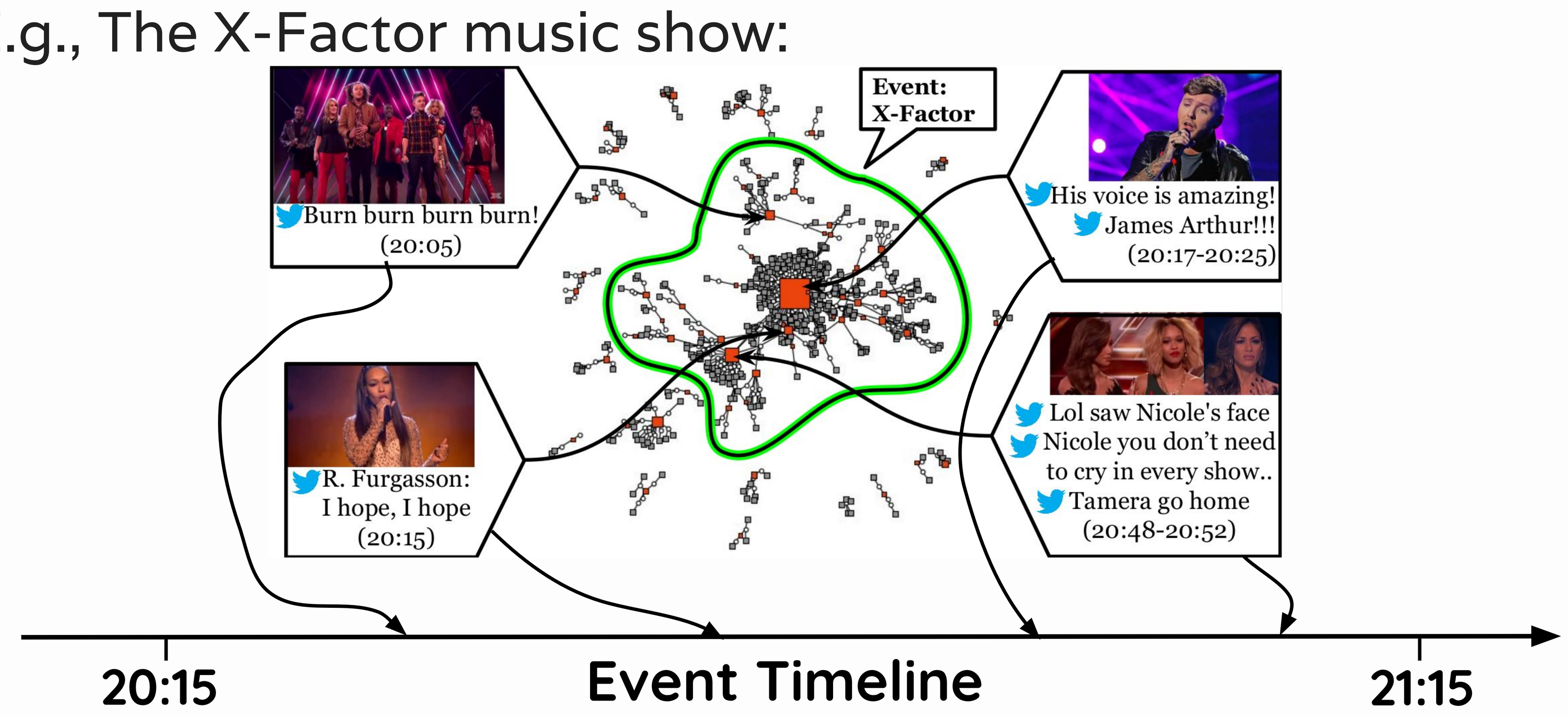
However, Event Detection techniques report and treat events as *singletons*

We introduce **Event Delineation** to better analyze, understand and narrate what happened

DELINEATION

Event Delineation is the breakdown of an event into its sub-events

E.g., The X-Factor music show:



DEFINITIONS

i) Snapshot of a Dynamic Network, $G_t = \{V_t, E_t\}$, with $V_t = V_{u,t} \cup V_{c,t}$, where $V_{u,t}$ is the set of user nodes and $V_{c,t}$ is the set of content nodes at time t, $E_t \subseteq V_t \times V_t$

ii) Event, $\Gamma = \{R_\Gamma, T_\Gamma, S_\Gamma\}$, where:

R_Γ: a representative summary (e.g., text description),

T_Γ: the time duration of the event, and

S_Γ: a set of sub-events

iii) Sub-Event, $\gamma_n = \{r_{\gamma\_n}, t_{\gamma\_n}\}$, where:

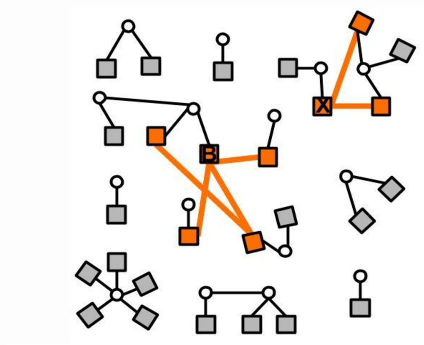
r_{γ_n}: representative summary, t_{γ_n}: timestamp of the sub-event

iv) Problem: Event Detection and Delineation

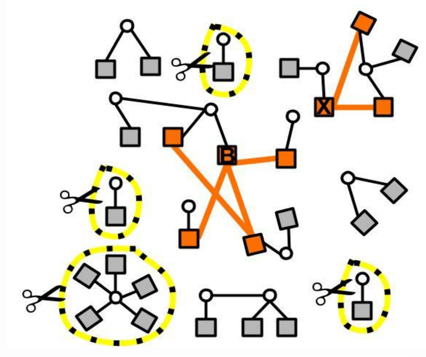
Given a Content Network G={G_t | t=0, ...}, find all events **Γ with all their sub-events** S_Γ

METHOD: DeLi

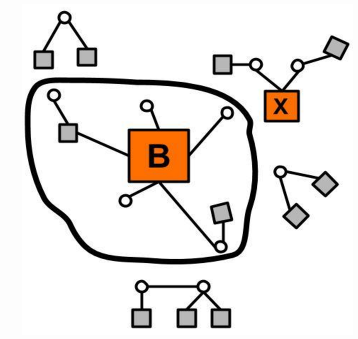
1) Build snapshot network & Reveal hidden links: *user*X replies *user*Y, *user*X posts *text*A, *text*A similar_to *text*B



2) Prune noise and spam: repetitive incidents & star structure CCs



3) Detect events and sub-events: events → very large CCs; sub-events → central content nodes in vlCCs



$$h(CC_i) = \begin{cases} 1, \text{ if } |CC_i| > avg(|CC|) + \theta * std(|CC|) \\ 0, \text{ otherwise} \end{cases}$$

EXPERIMENTS

1) Dataset: ~ 700K public geotagged **tweets** from London organised into 15-min time windows

Ground truth: Wikipedia & manual annotation

2) Event Detection

Act: unexpected #tweets

Struct: unexpected CC size based on network's structure (only user nodes)

$DeLi_{Con}$: unexpected CC size based on network's structure (only content nodes)

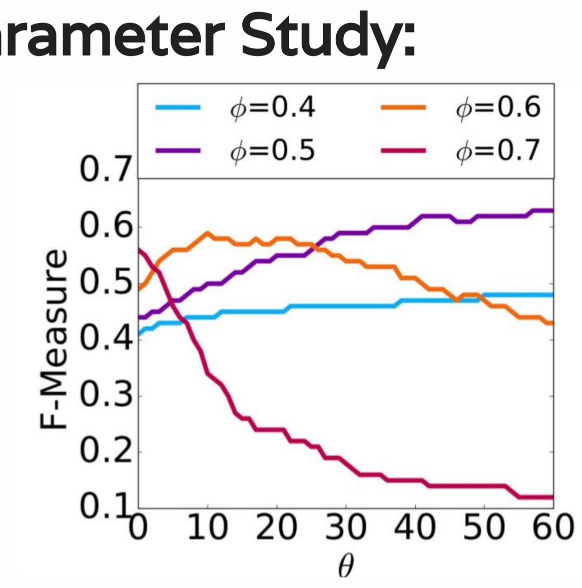
$DeLi_{Ter}$: unexpected CC size on the *Term Graph*

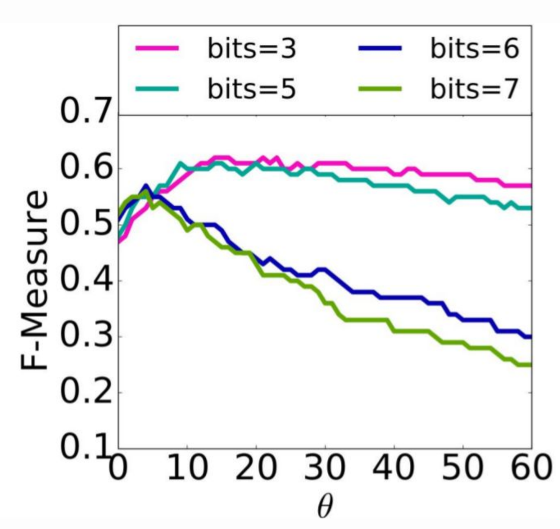
DeLi: using tfidf vectors for revealing similar links

DeLi#: using SimHash for revealing similar links

Method	Precision	Recall	F-Score
Act	0.33	0.70	0.45
Struct	0.28	0.87	0.42
K-Cores	0.21	0.43	0.28
$DeLi_{Con}$	0.44	**0.90**	0.59
$DeLi^{\#}_{Con}$	0.48	0.32	0.38
$DeLi_{Ter}$	0.39	0.95	0.55
DeLi	**0.53**	0.78	**0.63**
DeLi#	0.56	0.69	0.62

3) Parameter Study:





- Left: DeLi's F1-score for varied values of θ and φ (cosine similarity threshold)
- Right: DeLi#'s F1-score against θ and #bits (the SimHash fingerprint size)

4) Sub-Event Detection:

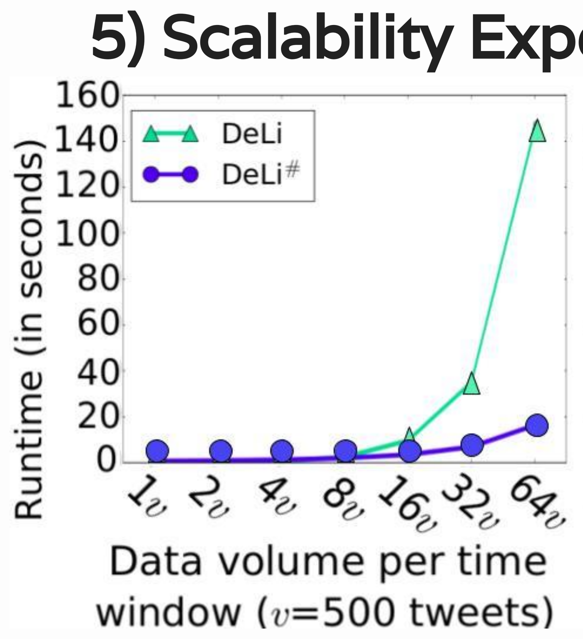
State-of-the-art: K-Cores utilizes the Term Graph and ranks events using the k-core score

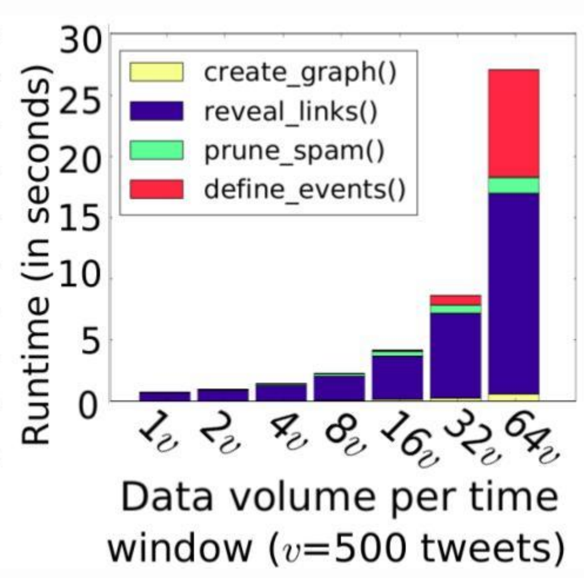
Method	Precision	Recall	F-Score	Runtime
K-Cores (1-min)	0.14	0.19	0.16	-
K-Cores (15-min)	0.2	0.03	0.05	11091s
DeLi	**0.15**	**0.49**	**0.22**	**139s**

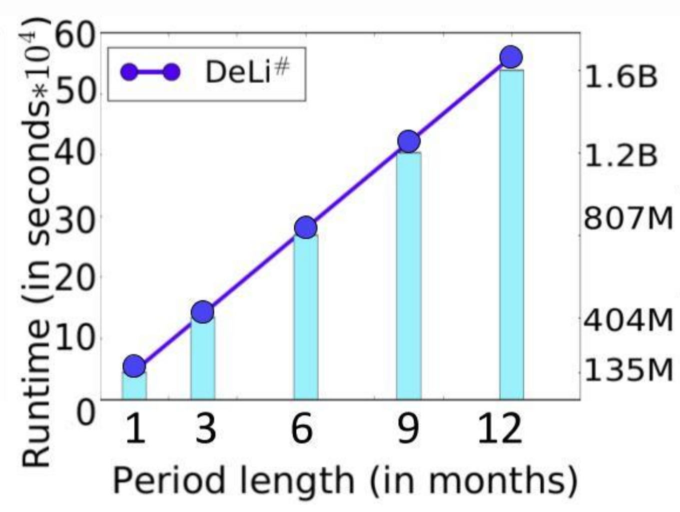
The fraction of sub-events of each type detected by each method

Sub-Event Type	Ground Truth	DeLi 15-min	K-Cores 1-min	K-Cores 15-min
Total goals	86	**0.70**	0.33	0.05
Penalties	3	**0.33**	0.0	0.0
Own goals	6	**0.83**	0.17	0.0
Yellow cards	85	**0.17**	0.01	0.0
Red cards	1	**1.0**	**1.0**	0.0
Final Score	29	**0.72**	0.31	0.07
All Sub-Events	210	**0.49**	0.19	0.03

5) Scalability Experiments:







- Left: Varying volume per time window
- Center: Execution time of each module
- Right: Varying time period (static volume per time window)

Acknowledgements: This research has been financed by the Horizon2020 VaVeL project (www.vavel-project.eu/), the FP7 ERC IDEAS NGHCS project (pages.cs.aueb.gr/~vana/NGHCS/), and a Google Faculty Research Award.

VaVeL NGHCS Google Faculty Research Awards

 contact: antoniasar@di.uoa.gr, www.di.uoa.gr/~antoniasar