Social Sensing for Urban Security

Salvatore Gaglio, Marco Morana

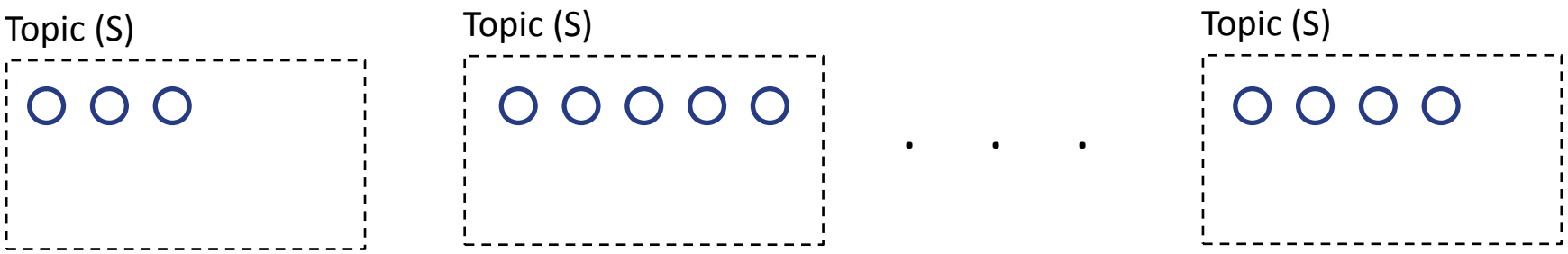
University of Palermo, Italy

Motivations and Goals

- A **a system for real-time Twitter data analysis** in order to follow a generic event from the user's point of view
  - Tweets are usually related to **events which involve many people** in different parts of the world
  - The framework we propose **adapts its behaviour to the nature of incoming data**

Find Best-Matching Terms

while (cosine similarity (S,t) > Θ)*
expand S;



Soft Frequent Pattern Mining

Soft Frequent Pattern Mining

```
function SFPM(C, K, b, c)
    T = SFPM_TermSelection(C, K);
    for each term t in T do
        Compute D_t;
    for each term t in T do
        S ← t; D_S ← D_T;
        expand ← true;
        repeat
            t* ← BestMatchingTerm(D_S, S, T);
            sim ← CosineSimilarity(D_S, D_t*)
            if (sim > Θ_{b,c}(S)) then
                S ← S ∪ t*;
                D_S ← D_s + D_t*;
                for i=1 to n do
                    if (D_Si < |S|/2) then
                        D_Si ← 0;
                    else
            else
                expand ← false;
        until expand
        Topics ← Topics ∪ S;
    return RemoveDuplicates(Topics);
```

S: *a set of terms (topic)*

D_S: *vector of i elements, stores how many of the terms in S co-occur in the i-th document*

D_t: *binary vector of i elements, where Dt(i) = 1 if the term t occurs in the i-th document.*

Dynamic Soft Frequent Pattern Mining

Soft Frequent Pattern Mining

```
function SFPM(C, K, b, c)
    T = SFPM_TermSelection(C, K);
    for each term t in T do
        Compute D_t;
    for each term t in T do
        S ← t; D_S ← D_T;
        expand ← true;
        repeat
            t* ← BestMatchingTerm(D_S, S, T);
            sim ← CosineSimilarity(D_S, D_t*)
            if (sim > Θ_{b,c}(S)) then
                S ← S ∪ t*;
                D_S ← D_s + D_t*;
                for i=1 to n do
                    if (D_Si < |S|/2) then
                        D_Si ← 0;
                    else
            else
                expand ← false;
        until expand
        Topics ← Topics ∪ S;
    return RemoveDuplicates(Topics);
```

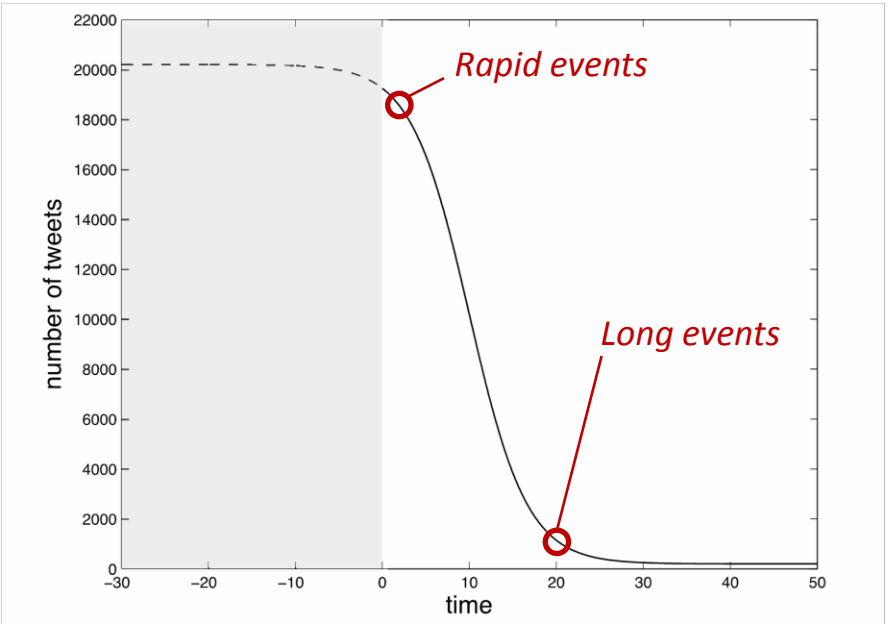
Term Selection

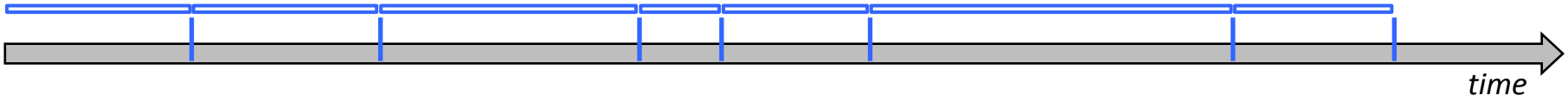
```
function TLDF_TermSelection(C, K)
    for each term t in C do
        p_new ← LikelihoodOfAppearance(t, C_new);
        p_ref ← LikelihoodOfAppearance(t, C_ref);
        r_t ← p_new / p_ref;
        TFIDF_t ← ComputeTFIDF(t);
        if (NER(t)) then
            ω_t ← 1.5;
        else
            ω_t ← 1;
        f_t ← ω_t × r_t × TFIDF_t;
    Sort(f, ASCENDING);
    for i=1 to K do
        T ← T ∪ t(f_i);
    return T
```

Dynamic Soft Frequent Pattern Mining

Dynamic Detection Windows

The size of the windows depends on the duration of the event and the number of related tweets





Dynamic Soft Frequent Pattern Mining

Dynamic Term Selection (DTS)

Select the terms which are relevant both in W_{n-1} (existing topics) and in W_n (emerging topics)

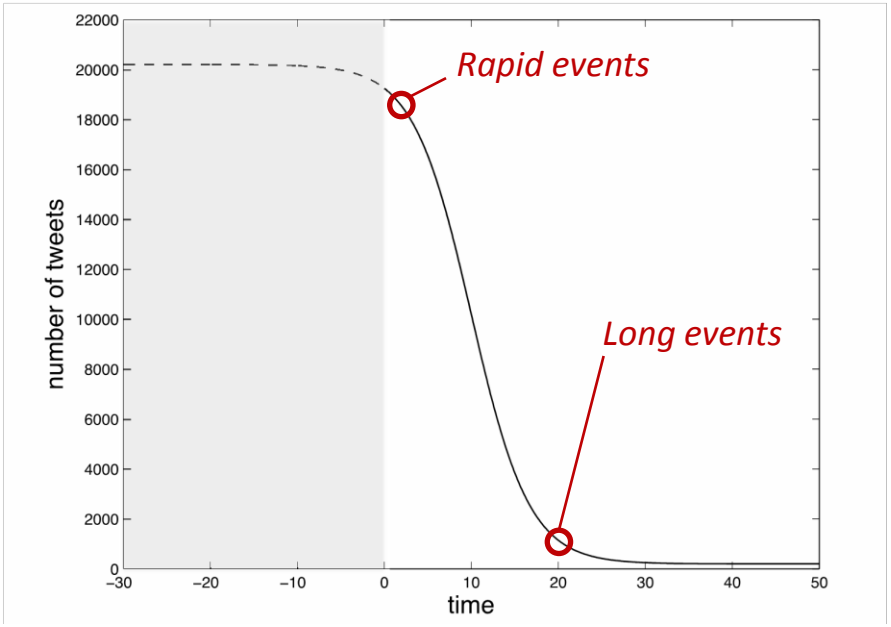
⇨ TF-IDF weighting

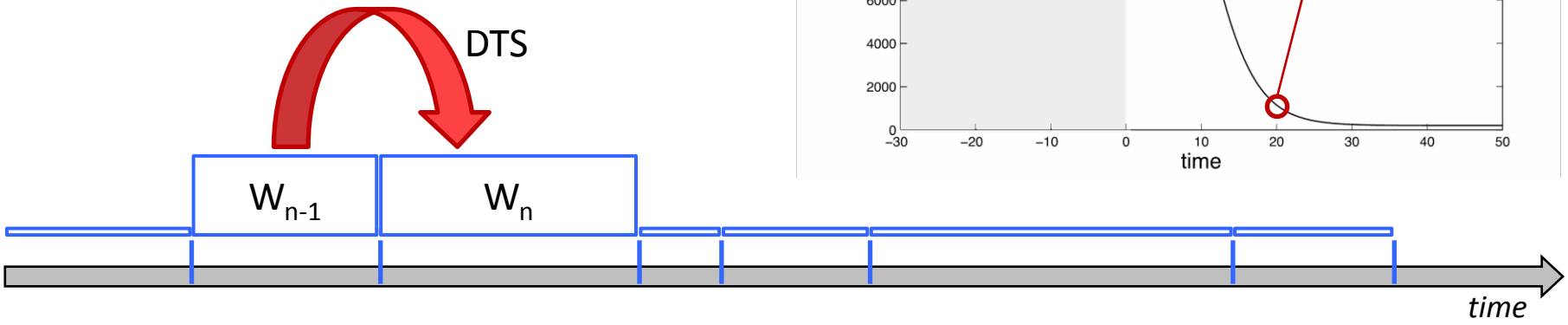
Give priority to terms denoting persons, organizations, and locations

⇨ Named-Entity Recognition (NER)

Dynamic Detection Windows

The size of the windows depends on the duration of the event and the number of related tweets





Dynamic Soft Frequent Pattern Mining

Dynamic Term Selection (DTS)

Select the terms which are relevant both in W_{n-1} (existing topics) and in W_n (emerging topics)

⇨ TF-IDF weighting

Give priority to terms denoting persons, organizations, and locations

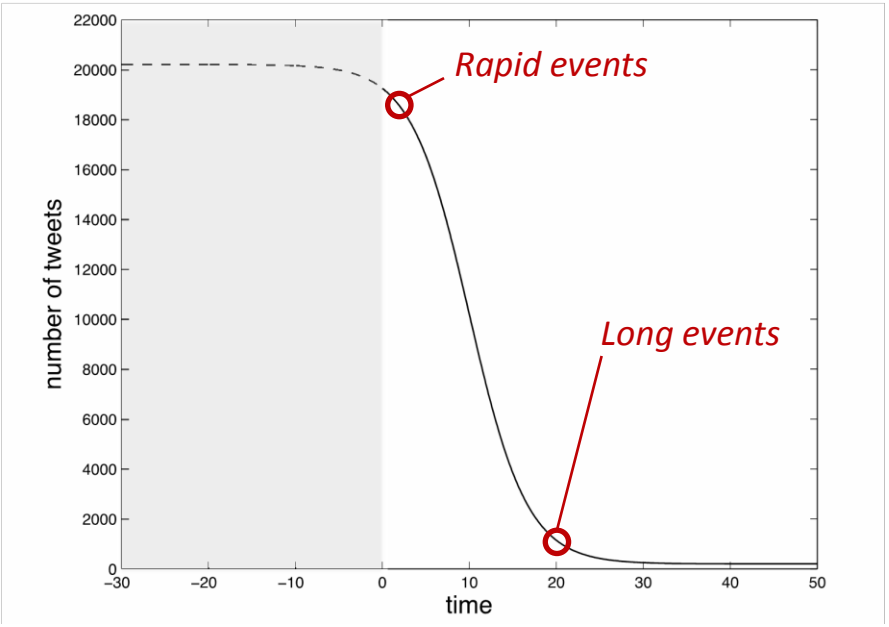
⇨ Named-Entity Recognition (NER)

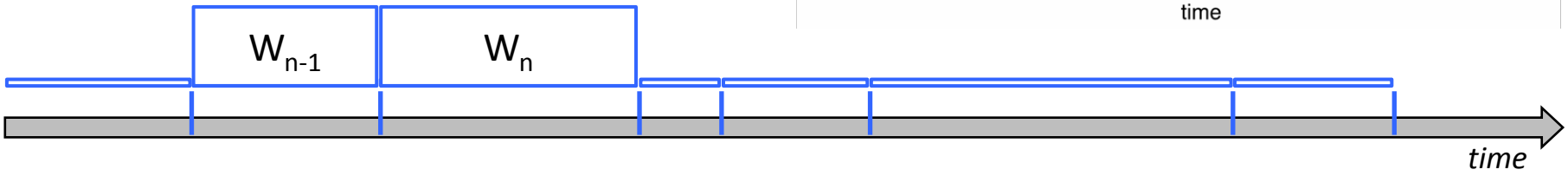
Dynamic Keywords

For each W_n the set of keywords is updated by including new terms which reflect the users' perspective on a specific event or deleting those unused.

Dynamic Detection Windows

The size of the windows depends on the duration of the event and the number of related tweets





Case Study

FIFA WORLD CUP Brasil

Case Study

FIFA WORLD CUP Brasil

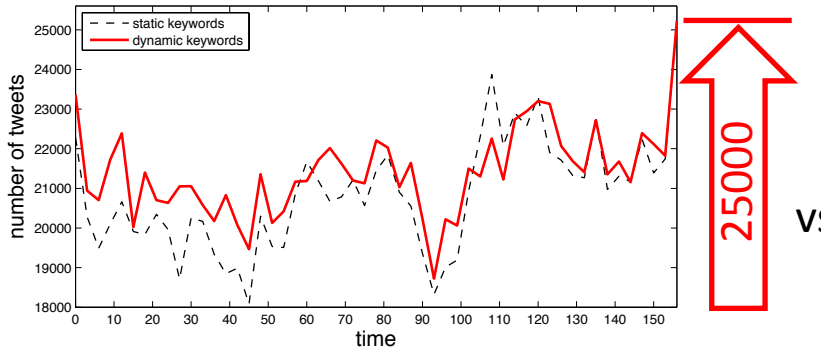
A framework for real-time Twitter data analysis. *S. Gaglio, G. Lo Re, M. Morana. In Journal of Computer Communications, Elsevier*

Real-Time Detection of Twitter Social Events from the User's Perspective. *S. Gaglio, G. Lo Re, M. Morana. In Proceedings of the 2015 IEEE International Conference on Communications (ICC2015)*

Results

GERMANY - ARGENTINA

#brasil2014, #brazil, #brasil, #worldcup2014, #worldcup, #FIFAWC2014, #ARGVsGER, #GERVsARG

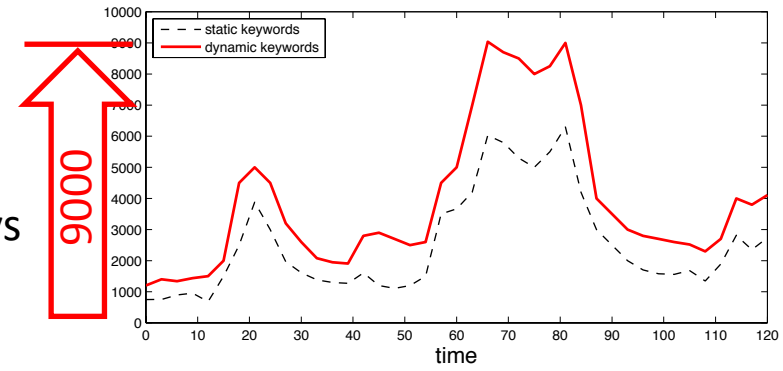


Average window duration: 2.3 min

NETHERLANDS - AUSTRALIA

#brasil2014, #brazil, #brasil, #worldcup2014, #worldcup, #FIFAWC2014, #NETvsAUS, #AUSvsNET

vs



Average window duration: 12.3 min

The 15 most popular topics detected during the FIFA World Cup 2014 final match

Time (CEST)	Topic
20:45	Puyol shows the World Cup trophy.
21:00	Khedira is out due to injury.
21:30	Higuain goal disallowed for offside.
21:47	Howedes's shot hits the near post.
22:30	Will the game be decided in 90 minutes?
22:53	Full-time! Match moves to extra-time.
23:00	Great chance for Schurrle.
23:14	First period of extra-time ends. Still 0-0.
23:17	*Messi was seen vomiting.*
23:24	GOAL! Mario Goetze.
23:36	Germany are the champions of the world.
23:36	Germany win the World Cup.
23:37	Gotze goal crowns Germany champions.
23:38	Germany has won its 4th title.
23:58	Germany lift the World Cup.

The 15 most popular topics detected during the FIFA World Cup 2014 final match

Time (CEST)	Topic
20:45	Puyol shows the World Cup trophy.
21:00	Khedira is out due to injury.
21:30	Higuain goal disallowed for offside.
21:47	Howedes's shot hits the near post.
22:30	**Will the game be decided in 90 minutes?**
22:53	Full-time! Match moves to extra-time.
23:00	Great chance for Schurrle.
23:14	First period of extra-time ends. Still 0-0.
23:17	***Messi was seen vomiting.***
23:24	GOAL! Mario Goetze.
23:36	Germany are the champions of the world.
23:36	Germany win the World Cup.
23:37	Gotze goal crowns Germany champions.
23:38	**Germany has won its 4th title.**
23:58	Germany lift the World Cup.

The 15 most popular topics detected during the FIFA World Cup 2014 final match

Event? Spam? or… user's perspective?

Time (CEST)	Topic
20:45	Puyol shows the World Cup trophy.
21:00	Khedira is out due to injury.
21:30	Higuain goal disallowed for offside.
21:47	Howedes's shot hits the near post.
22:30	**Will the game be decided in 90 minutes?**
22:53	Full-time! Match moves to extra-time.
23:00	Great chance for Schurrle.
23:14	First period of extra-time ends. Still 0-0.
23:17	***Messi was seen vomiting.***
23:24	GOAL! Mario Goetze.
23:36	Germany are the champions of the world.
23:36	Germany win the World Cup.
23:37	Gotze goal crowns Germany champions.
23:38	**Germany has won its 4th title.**
23:58	Germany lift the World Cup.

Social Sensing for Urban Security

We can process social networks data to detect relevant events related to a given topic of interest

- The users of the social networks are citizens first
- In a smart city citizens can be used as sensors which provide real-time information about the city and its quality of living
- The social networks can be considered as a part of a heterogeneous sensing infrastructure which also includes hardware sensors, whether wearable or embedded in mobile devices.
- An ICT platform can be designed to discover implicit information contained in social posts
- Topic detection and summarization techniques can be used for automatic reporting of significant events

UNIVERSITÀ
DEGLI STUDI
DI PALERMO

http://www.dicgim.unipa.it/networks

References

1. A framework for real-time Twitter data analysis. Gaglio, G. Lo Re, M. Morana. In Journal of Computer Communications, ISSN 0140-3664. doi: /10.1016/j.comcom.2015.09.021.

2. Real-Time Detection of Twitter Social Events from the User's Perspective. Gaglio, G. Lo Re, M. Morana. In Communications (ICC), 2015 IEEE International Conference on, pp.1207-1212, 8-12 June 2015. doi: 10.1109/ICC.2015.7248487