

# Online Centrality in Temporally Evolving Networks

Ferenc Béres and András A. Benczúr

Informatics Laboratory of the Hungarian Academy of Sciences (MTA SZTAKI)

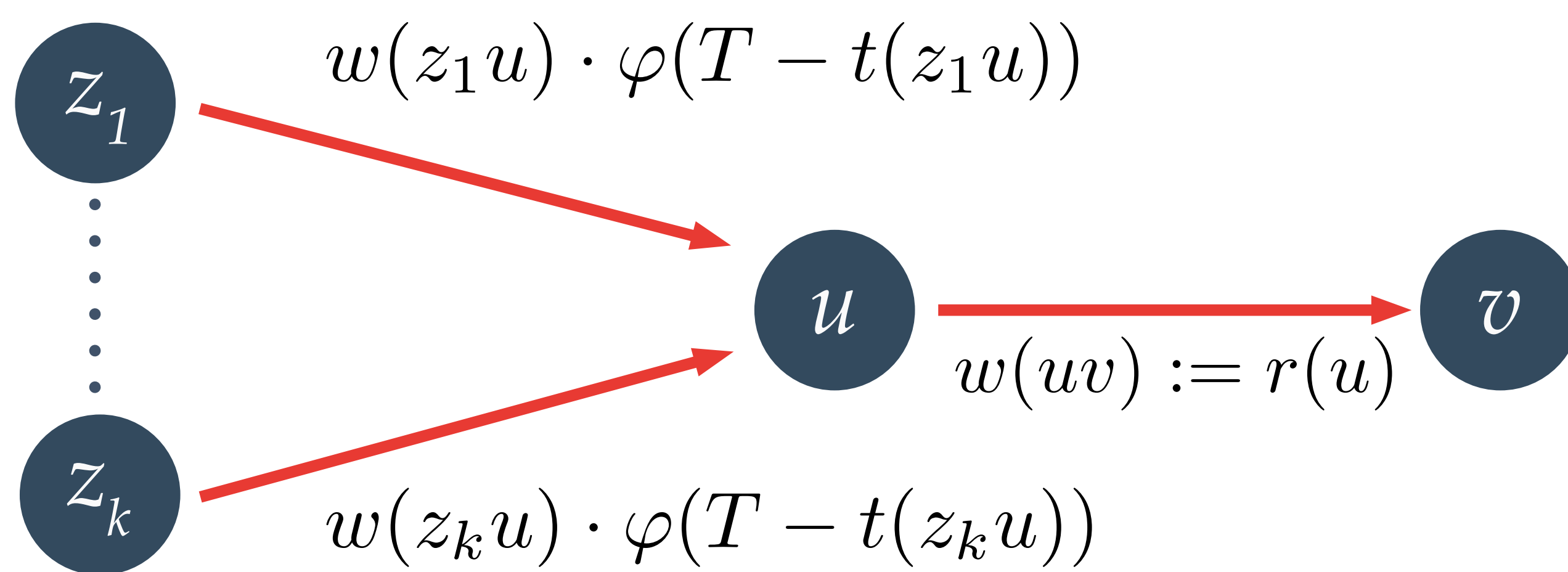
## OBJECTIVE

- We introduce a centrality measure updateable by the edge stream in a dynamic network
- It incorporates the elapsed time of edge activations

## ONLINE CENTRALITY

- $D = (V, A)$ : dynamic directed graph, e.g.: mention network
- Edges can be activated multiple times, cannot be deleted
- $A(T)$ : set of edges activated before time  $T$ ,  $A(0)=\emptyset$ .
- $w(uv)$ : weight over edge  $uv$
- $t(uv)$ : last edge activation time of edge  $uv$
- $\varphi(x)$ : time decay function, vanishes in infinity
- $r(v)$ : the Online Centrality score of node  $v$

## UPDATE RULE



- Edge  $uv$  activated at time  $T$ , we update: (1)  $r$ , (2)  $w$  and  $t$

$$r(u) := \alpha + \sum_{zu \in A(T)} w(zu) \cdot \varphi(T - t(zu))$$

$$w(uv) := r(u)$$

$$t(uv) := T$$

## TIME DECAY

- Decay intensity is controlled by parameter  $n$
- $\Delta t$  is measured in seconds
- Exponential decay:  $\varphi(\Delta t) := b^{\frac{\Delta t}{n}}, (0 < b < 1)$
- Rayleigh decay:  $\varphi(\Delta t) := \frac{1}{\sigma^2} \cdot \frac{\Delta t}{n} \cdot e^{-\frac{1}{2\sigma^2} \cdot (\frac{\Delta t}{n})^2}$

By the end of the day most of the models can predict daily tennis players accurately

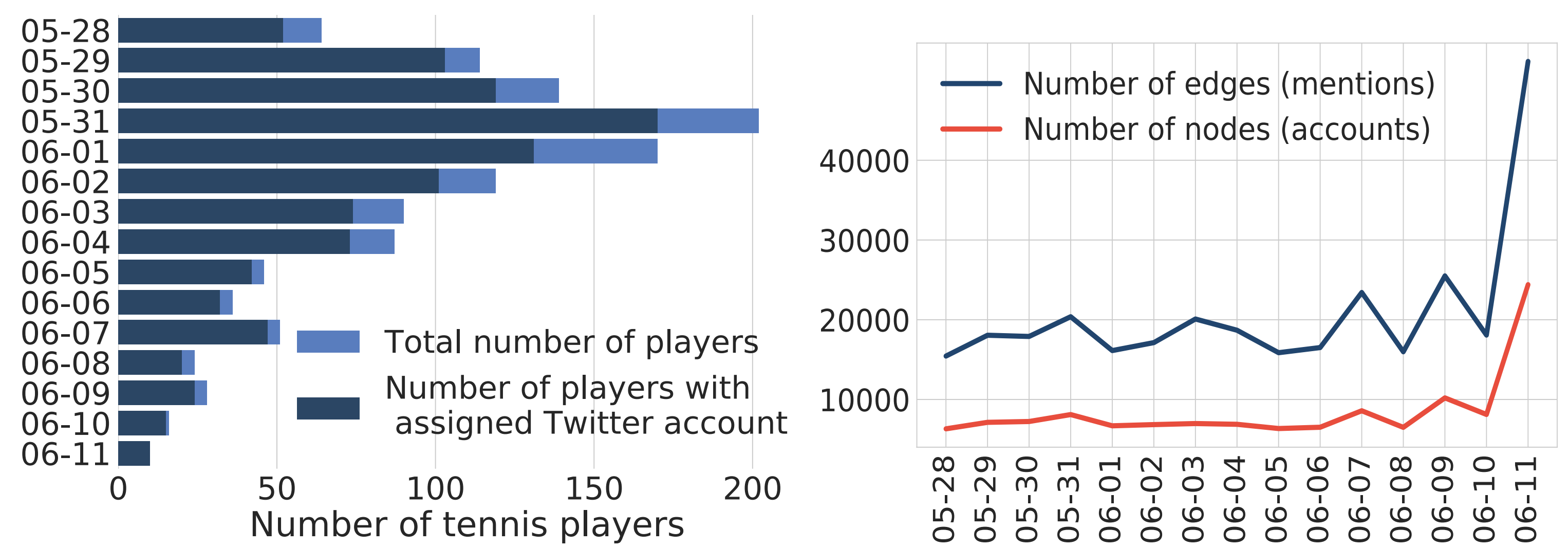
Online Centrality with both Exponential (*online-Exp*) and Rayleigh time decay (*online-Ray*) outperforms baselines for all snapshots

Snapshot-12 methods are slow at detection, Snapshot-1 methods give less increase and decline too early

Our methods are best at early detection, from 2AM to 2PM, before the actual games.

## ROLAND-GARROS DATASET

- We collected tweets from May 23 to June 16 using keywords: *@rolandgarros, #rolandgarros2017, #frenchopen, #rg17*
- 444,328 tweets, 351,692 mentions
- Mention network: nodes are accounts, edges are mentions
- Assign tennis players to Twitter accounts (412 account found)
- Filter for relevant players: Men's, Women's, Legends Under 45
- Tournament days from 2017 May 28 to June 11 (15 days)



## MEASUREMENT

- Central node prediction in the temporally evolving mention network of Roland-Garros 2017
- **Goal:** predict list of players who play on a given day
- **Evaluation** by 
$$\text{NDCG} = \frac{1}{\text{IDCG}} \sum_{i=1}^{\infty} \frac{\text{rel}(i)}{\log_2(i+1)},$$
  - ranking is based on network centrality
  - $\text{rel}(i) = 1$  if node  $i$  is related to a player who participated on the given day and 0 otherwise
- **Baselines:** Temporal PageRank [Rozenshtein et al.] (**temp-PR**) and static centrality measures; PageRank (**PR**), Indegree (**indeg**), Negative-Beta measure (**NBM**), Harmonic centrality (**HC**); calculated in 1 or 12 hour time windows
- NDCG averages of 15 days, with confidence intervals shown

