Temporal Walk Based Centrality Metric for Graph Streams

Ferenc Béres
Eötvös Loránd University

Róbert Pálovics
Stanford University

András A. Benczúr
Hungarian Academy of Sciences

DYNAMIC GRAPHS

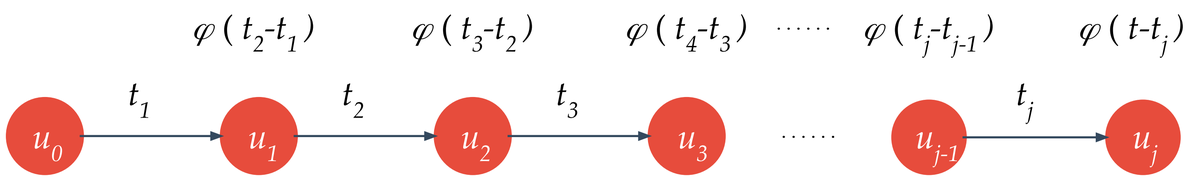
- Edge stream: (u,v,t) triplets, link occured from u to v at time t
- Directed edges can be activated multiple times, no deletion

TEMPORAL KATZ CENTRALITY

- Centrality measure for dynamic graphs
- Extension of Katz-index to temporally changing networks
- Online updateable from the edge stream

DEFINITION

- $\varphi(\tau)$: time decay function, e.g. $\varphi(\tau) = \beta\, exp(-c\tau)$
- Time-respective walk: $t_1 < t_2 < t_3 < \ldots < t_j \leq t$



$$\Phi(z,t) := \prod_{i=1}^{j} \varphi(t_{i+1} - t_i);\ t_{j+1} = t$$

- $r_u(t)$: temporal Katz centrality score of node u at time t
- Weighted sum of time-respective walks ending at node u

$$r_u(t) := \sum_{v} \sum_{\text{temporal paths } z \text{ from } v \text{ to } u} \Phi(z,t)$$

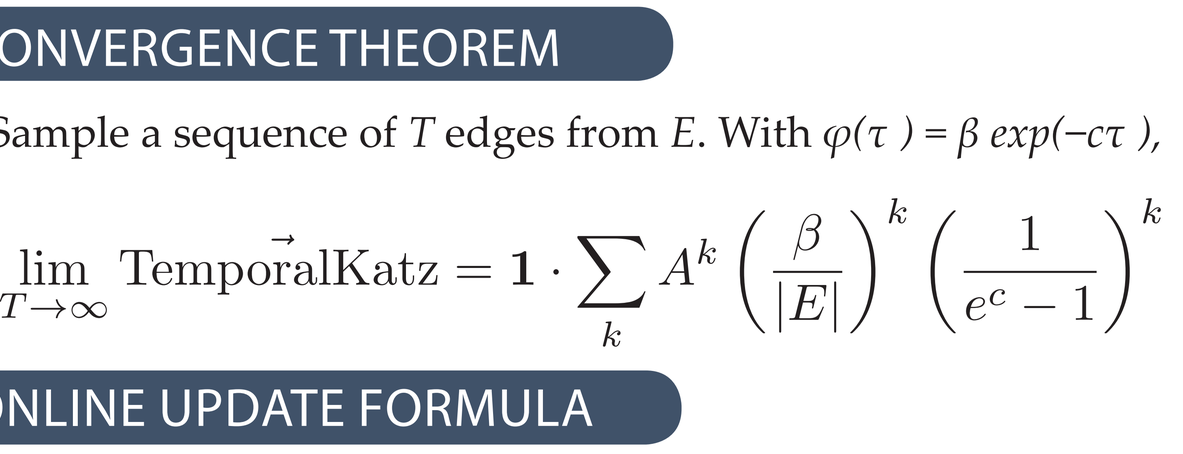
CONVERGENCE THEOREM

Sample a sequence of T edges from E. With $\varphi(\tau) = \beta\, exp(-c\tau)$,

$$\lim_{T\to\infty} \vec{\text{TemporalKatz}} = \mathbf{1} \cdot \sum_k A^k \left(\frac{\beta}{|E|}\right)^k \left(\frac{1}{e^c - 1}\right)^k$$

ONLINE UPDATE FORMULA

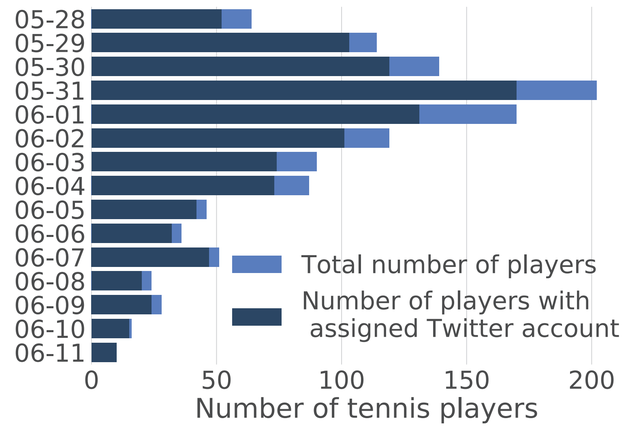
- $E(t)$: multi set of edges activated before time t
- r_v : temporal Katz centrality score of node v
- w_{vu} : weight over edge vu
- t_{vu} : last edge activation time of edge vu
- r,w and t are online updated from the edge stream



- More efficient update if $\varphi(\tau_1 + \tau_2) = \varphi(\tau_1)\,\varphi(\tau_2)$ e.g. exponential

TWITTER DATASETS

- Tweets from Roland-Garros and US Open 2017
- Mention network: nodes are accounts, edges are mentions
  - RG17: 336,234 edges; 78,095 nodes; 19 days
  - UO17: 482,061 edges; 106,920 nodes; 21 days
- We assigned tennis players to Twitter accounts
- Nodes of the network are temporally labeled
  - e.g. RG17 semi final day (June 9) at 12:00



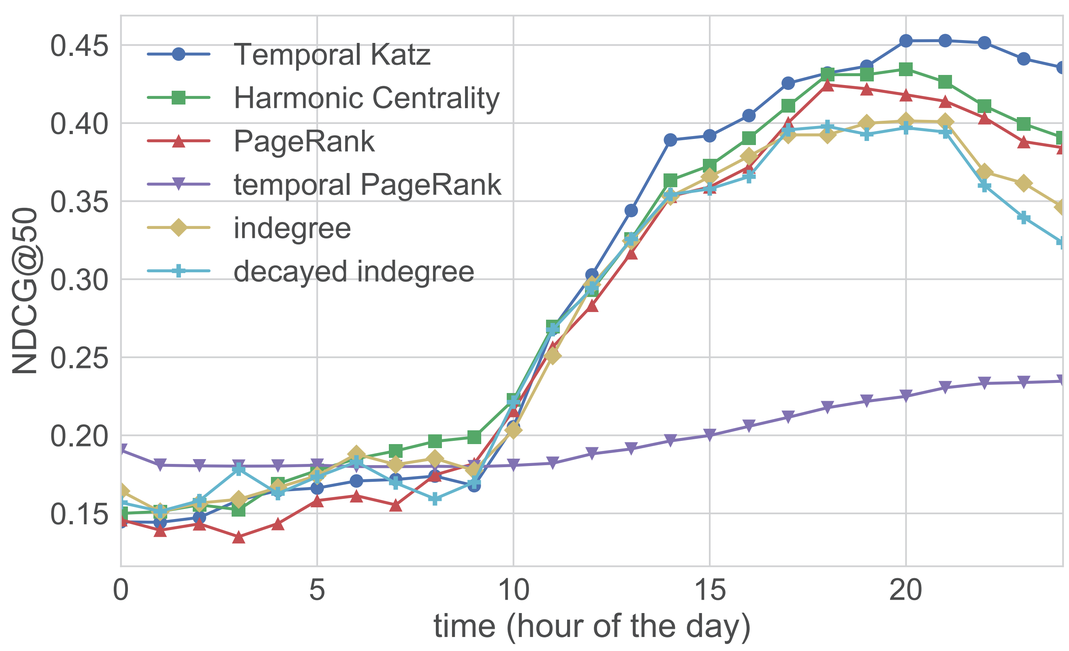
Rank	Node (Twitter account)	Label
1	@rolandgarros	0
2	@stanwawrinka	1
3	@andy_murray	1
4	@Simona_Halep	0
5	@RafaelNadal	1
6	@ThiemDomi	1
7	@TimeaOfficial	0
8	@AnaIvanovic	0
9	@rohanbopanna	0
10	@WTA	0

MEASUREMENT

- Central node prediction in temporally evolving networks
- **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
- rel(i) = 1 if node i is related to a player who participated on the given day and *0* otherwise



- Temporal Katz centrality outperforms baselines on both data

NDCG@50	UO17	RG17
indegree	0.321	0.342
decayed indegree	0.321	0.346
negative beta measure	0.319	0.333
PageRank	0.325	0.349
temporal PageRank	0.187	0.195
harmonic centrality	0.353	0.359
temporal Katz centrality	0.370	0.368

MTA SZTAKI
Hungarian Academy of Sciences
Institute for Computer Science and Control

beres@sztaki.hu, palovics@stanford.edu, benczur@sztaki.hu

All datasets and results are available on GitHub
https://github.com/ferencberes/online-centrality