Modeling Collective Anticipation and Response on Wikipedia

Renaud Lambiotte
Mathematical Institute
University of Oxford

Kobayashi, Ryota, et al. "Modeling collective anticipation and response on wikipedia." *arXiv preprint arXiv:2105.10900* (2021).

"I think the next century will be the century of complexity."

Stephen Hawking

com.plex
[adj., v. kuh m-pleks, kom-pleks; n. kom-pleks]

1) composed of many interconnected parts; compound; composite: a complex highway system

2) characterized by a very complicated or involved arrangement of parts, units, etc.: complex machinery

3) so complicated or intricate as to be hard to understand or deal with: a complex problem

Source: Dictionary.com

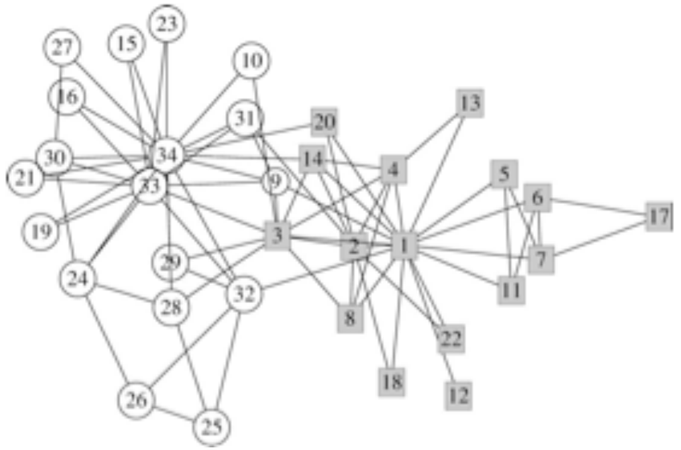
Box 1.1

"Complex systems consist of a large number of interacting components. The interactions give rise to emergent hierarchical structures. The components of the system and properties at systems level typically change with time."

H.J. Jensen, in Encyclopedia of Complexity and Systems Science

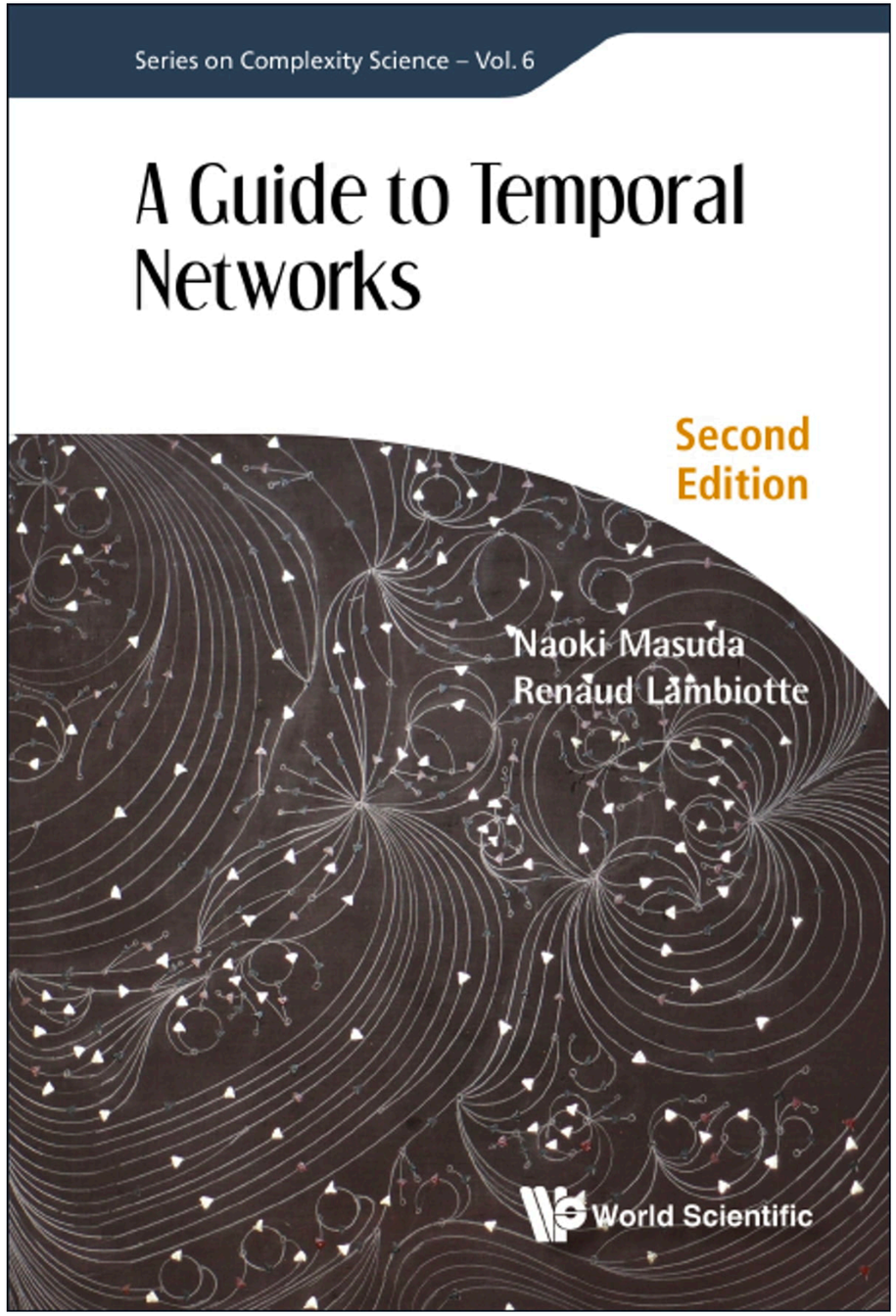
Tool box

Agent-based and numerical simulations
Information theory
(Non-linear) Dynamical systems
Data mining and optimisation
Networks



Social networks
- Collaboration networks
- Communication networks
- Online social networks

Biological networks
- Protein-protein interaction networks
- Neural networks
- Food webs



How does the structure of a network affects dynamics ?

How can we use dynamics to reveal important structural features of a network?

How does the structure of a network evolve in time?

How does a system repond to a perturbation?

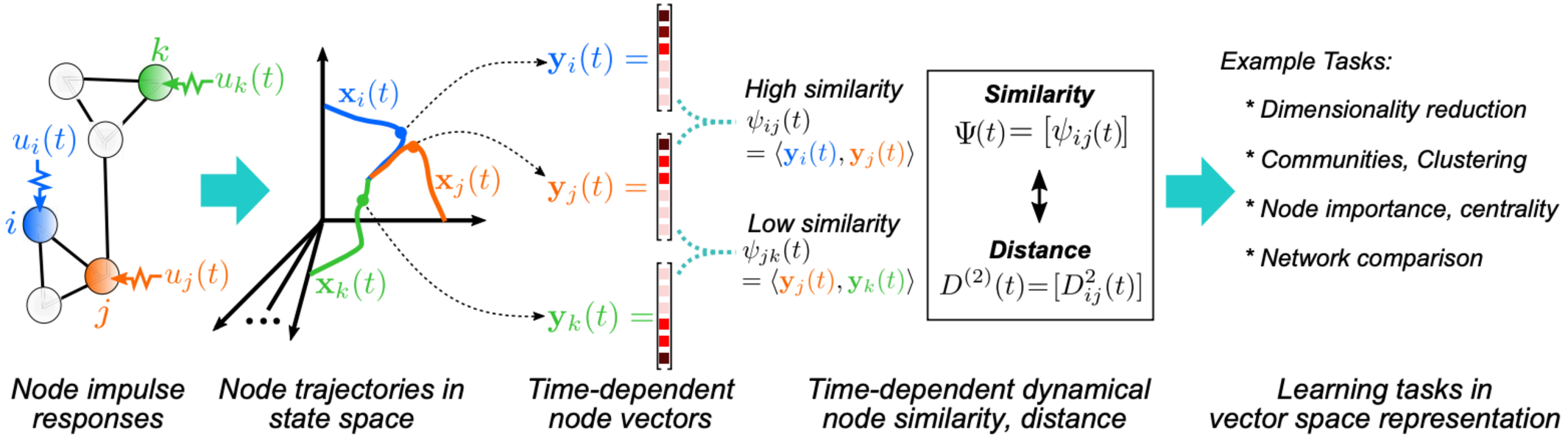
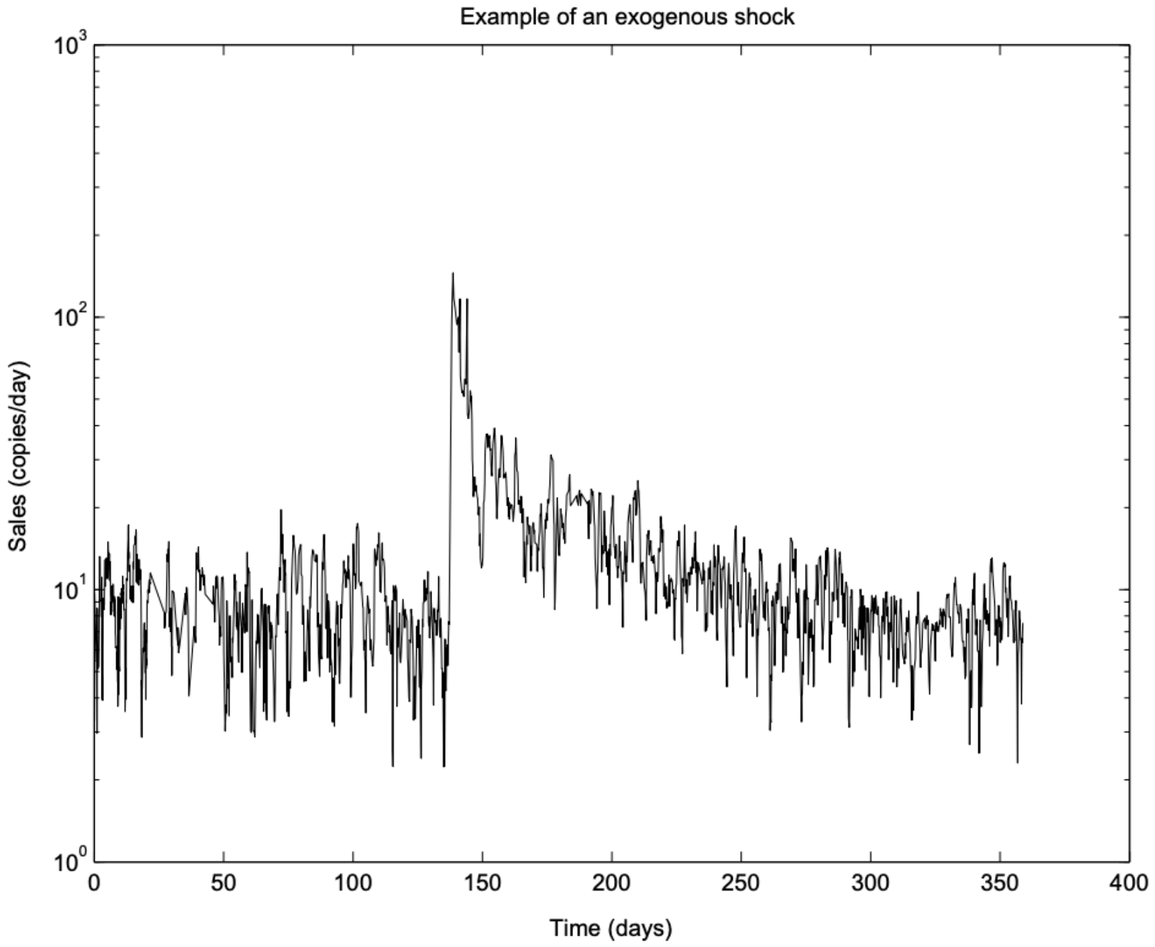


FIG. 1. Schematic of constructing dynamical similarity measures. Impulses are applied as inputs to different nodes of the network. The responses in time are interpreted as node vectors evolving in state space and can be compared, e.g., via an inner product from which we construct the similarity matrix $\Psi(t)$, or alternatively, its associated distance $D^{(2)}$. Nodes that drive the system similarly (differently) within the projected subspace are assigned a high (low) similarity score. The thus derived vector-space representation of the nodes can be used for a number of different learning tasks.

How does a system repond to a perturbation?



Interaction between agents

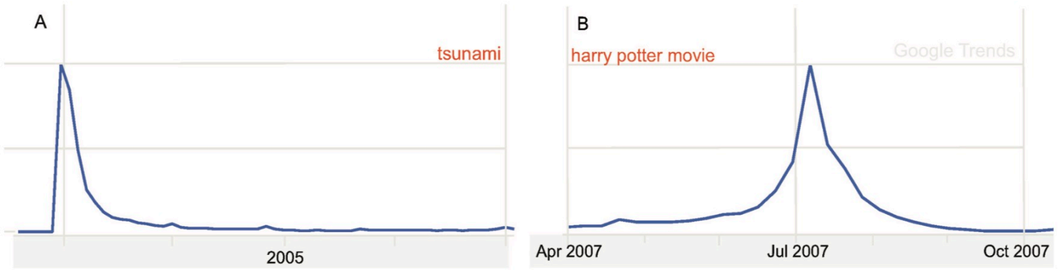
Self-exiting processes (e.g. Hawkes processes)

External perturbations

Sornette, Didier, et al. "Endogenous versus exogenous shocks in complex networks: An empirical test using book sale rankings." *Physical Review Letters* 93.22 (2004): 228701.

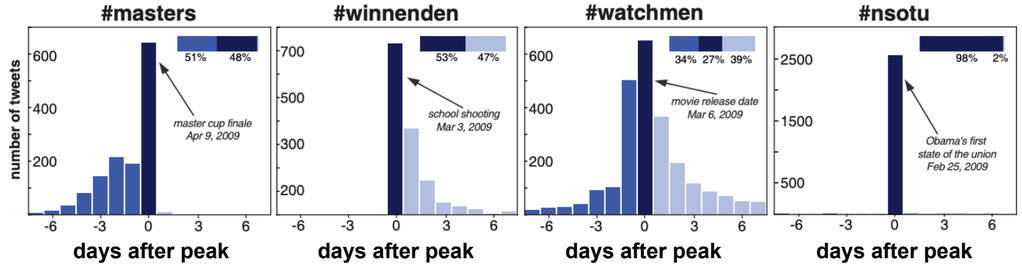
Peaks of activity in online media

Endogenous versus Exogeneous



Crane and Sornette. PNAS 2008

Classification into different classes



Lehmann et al. WWW 2012

Most of the research has focused on an unexpected shock, but what about a planned event,

e.g., Election, Football match, Film release,

Data

- We collected the time series of Wikipedia page views for pre-planned events.
- We consider 3,273 events of five popular categories:
  1. Election, 2. Sports event, 3. Association Football match (Football),
  4. Film release (Film), and 5. Holiday.
- We collected hourly page views for 10 days before and after the event day (21 days).

Example: 2016 NHL Winter Classic (Jan. 1st, 2016)

From Dec. 22nd 2015 to Jan. 11th, 2016

2016 NHL Winter Classic

From Wikipedia, the free encyclopedia

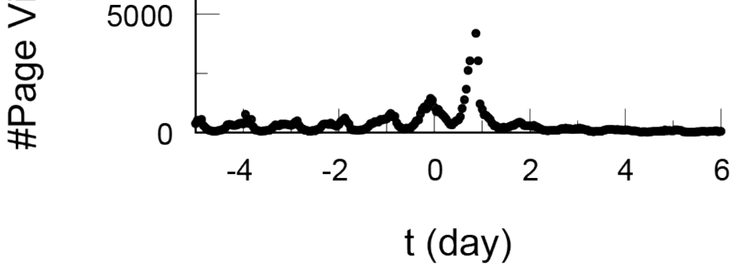
The **2016 NHL Winter Classic** (officially the **2016 Bridgestone NHL Winter Classic**)[1] was an outdoor regular season National Hockey League (NHL) game, part of the Winter Classic series, played on January 1, 2016. The game, the eighth Winter Classic, saw the Montreal Canadiens defeat the Boston Bruins, 5–1, at Gillette Stadium in Foxborough, Massachusetts, a significant event in one of the NHL's best-known rivalries. A Bruins and Canadiens alumni game was also played on December 31, 2015.[2] The Boston Pride women's professional hockey team played before the alumni game against Les Canadiennes of the Canadian Women's Hockey League to a 1–1 tie in the first ever 2016 Outdoor Women's Classic (officially the Outdoor Women's Classic presented

2016 NHL Winter Classic

	1	2	3	Total
Montreal Canadiens	1	2	2	5
Boston Bruins	0	0	1	1

Date January 1, 2016

Attention



https://en.wikipedia.org/wiki/2016_NHL_Winter_Classic

Research Questions (RQs)

We explore two research questions about temporal patterns of collective attention towards a planned event.

RQ1. What are the essential characteristics of the temporal pattern towards planned events?

RQ2. How is event-related information (e.g., category and outcome) associated with the temporal pattern?

Model

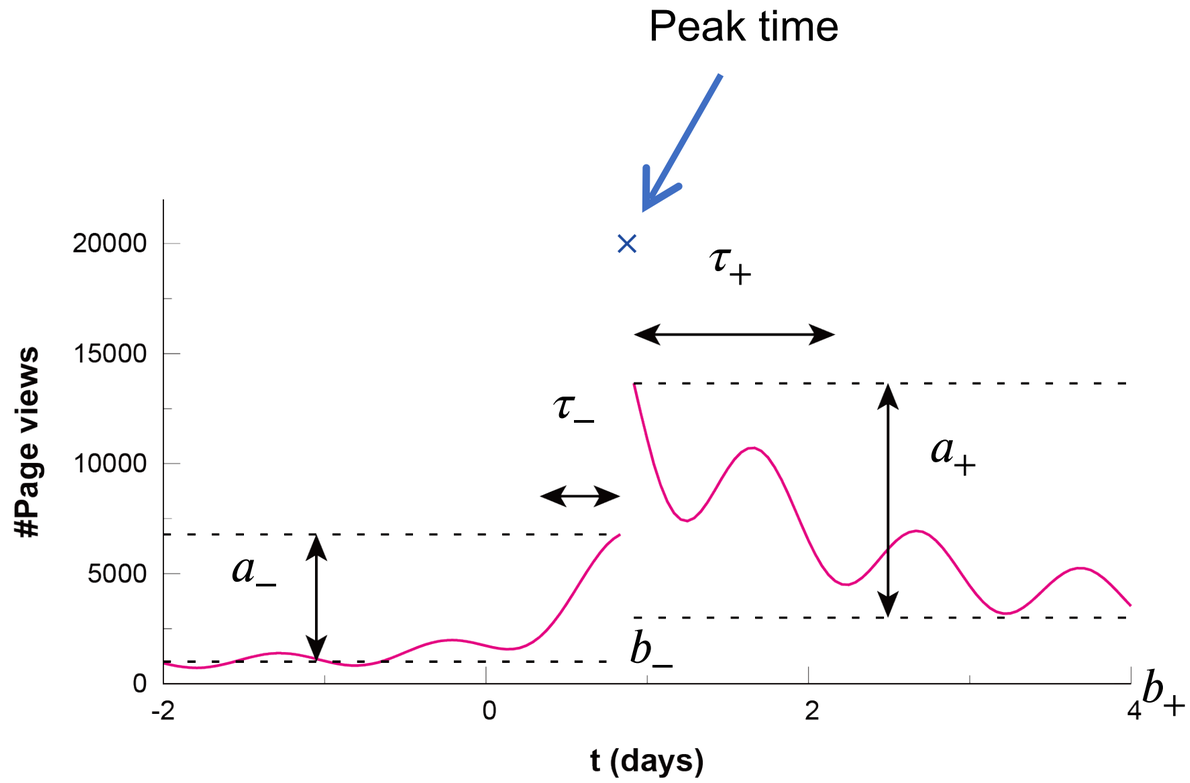
We develop a simple model for the attention dynamics that incorporates three characteristics,

1) Anticipation (before the event),
2) Response (after the event),
3) Circadian Rhythm.

$$f_{\text{peak}}(t) = C(t) D_{\text{peak}}(t)$$

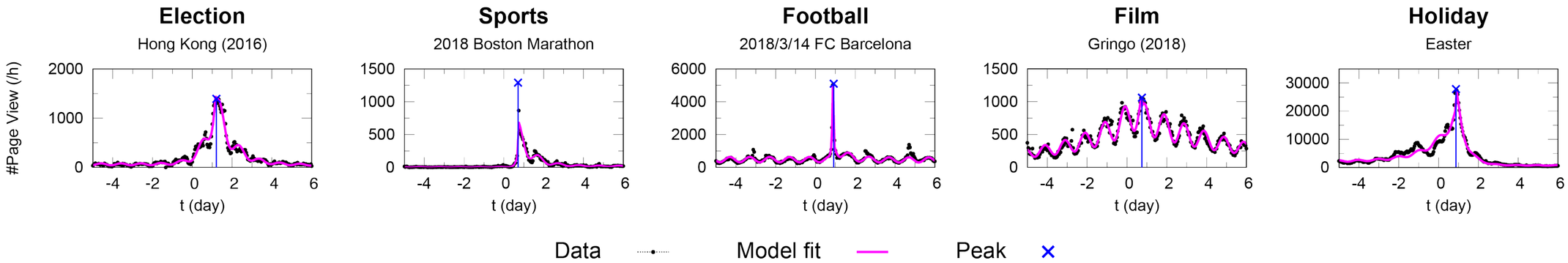
where $$C(t) = 1 + \alpha_c \cos\left(\frac{2\pi}{T}(t - t_c)\right)$$

$$D_{\text{peak}}(t) = \begin{cases} a_{-} e^{-\left|t - t_p\right|/\tau_{-}} + b_{-} & \left(t < t_p\right) \\ a_{+} e^{-\left|t - t_p\right|/\tau_{+}} + b_{+} & \left(t > t_p\right) \end{cases}$$



Model Fitting

The model can fit the attention dynamics accurately.



Peak Width	Broad	Moderate	Narrow	Broad	Broad
	Large: $\tau_-,\ \tau_+$		Small: $\tau_-,\ \tau_+$	Large: $\tau_-,\ \tau_+$	Large: $\tau_-,\ \tau_+$

The time constants are related to the event category.

Research Questions (RQs)

We explore two research questions about temporal patterns of collective attention towards a planned event.

RQ1. What are the essential characteristics of the temporal pattern towards planned events?

RQ2. How is event-related information (e.g., category and outcome) associated with the temporal pattern?

Clustering Attention Patterns

Clustering result based on the model parameters $\{a_{\pm},\ b_{\pm},\ \tau_{\pm};\ \alpha_c,\ t_c\}$ are associated with the event categories (Gaussian mixture model + BIC)

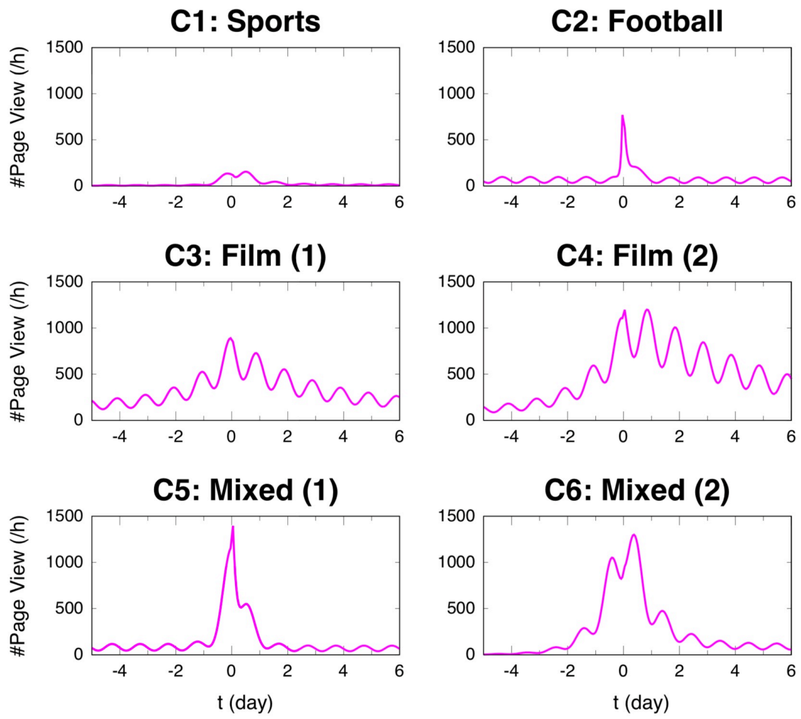


Figure: Cluster centers

Table: Identified clusters

	C1	C2	C3	C4	C5	C6
Election	3	2	4	3	16	64
Sports	**120**	7	1	4	11	70
Football	0	**165**	0	0	5	63
Film	0	0	**140**	**57**	26	6
Holiday	0	0	0	0	14	44

Table: Performance (AMI)

Method	Clusters	Features	AMI
Proposed	6	8	**0.47 (0.46–0.49)**
SpikeM	8	7	0.35 (0.33–0.38)
Power-law	7	4	0.36 (0.32–0.38)
Fraction	2	3	0.39 (0.39–0.39)

Predicting Response Patterns

We predict the page view time series for six days after the event date. We use a Bayesian method for fitting the response parameters, which utilizes the attention dynamics before the peak and event category information.

1. Setup

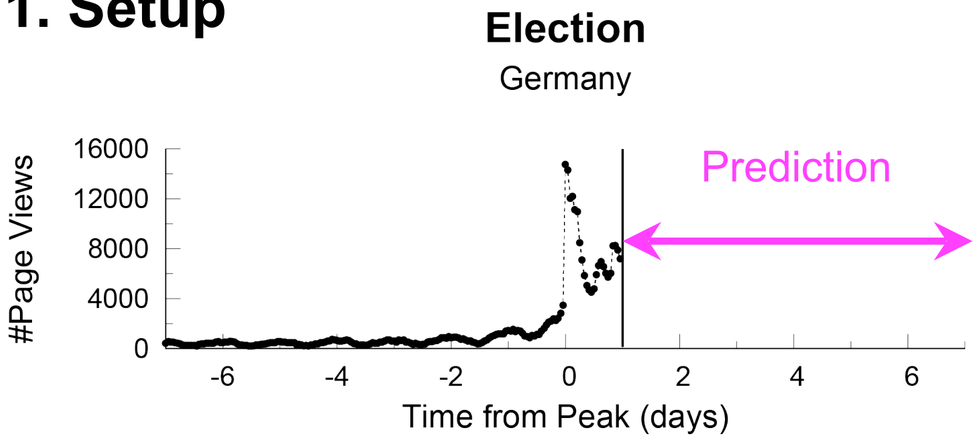
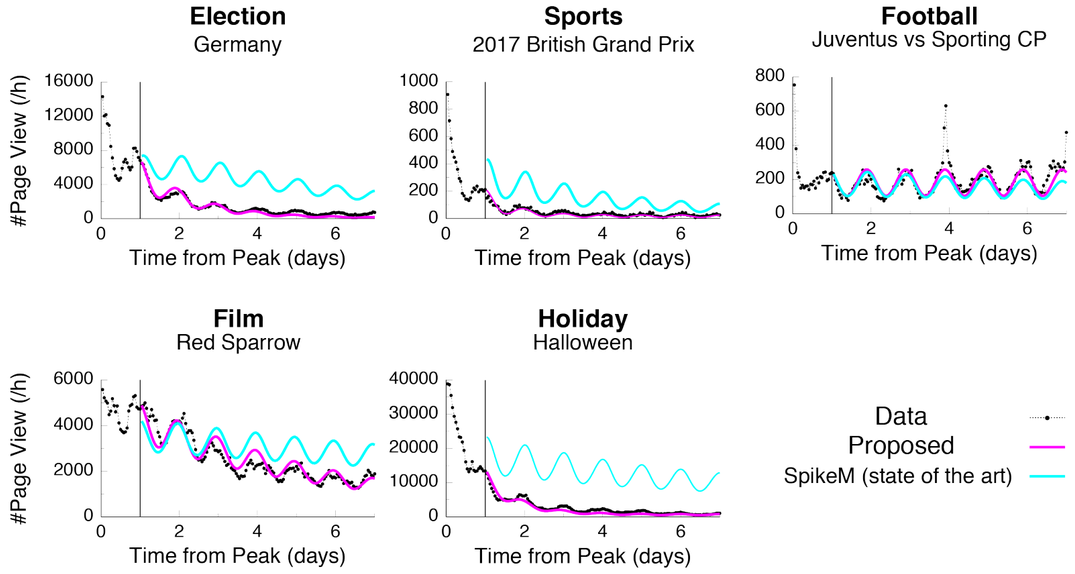


Table: Prediction error

Method	Complexity	1 day	2 days	3 days
Proposed	8	**0.54**	**0.51**	**0.46**
SpikeM	7	0.81	0.68	0.59
Power-law	4	0.68	0.64	0.62
LR	≥ 192	0.69	0.56	0.54
LSTM	$\approx 160{,}000$	1.4	1.0	0.83

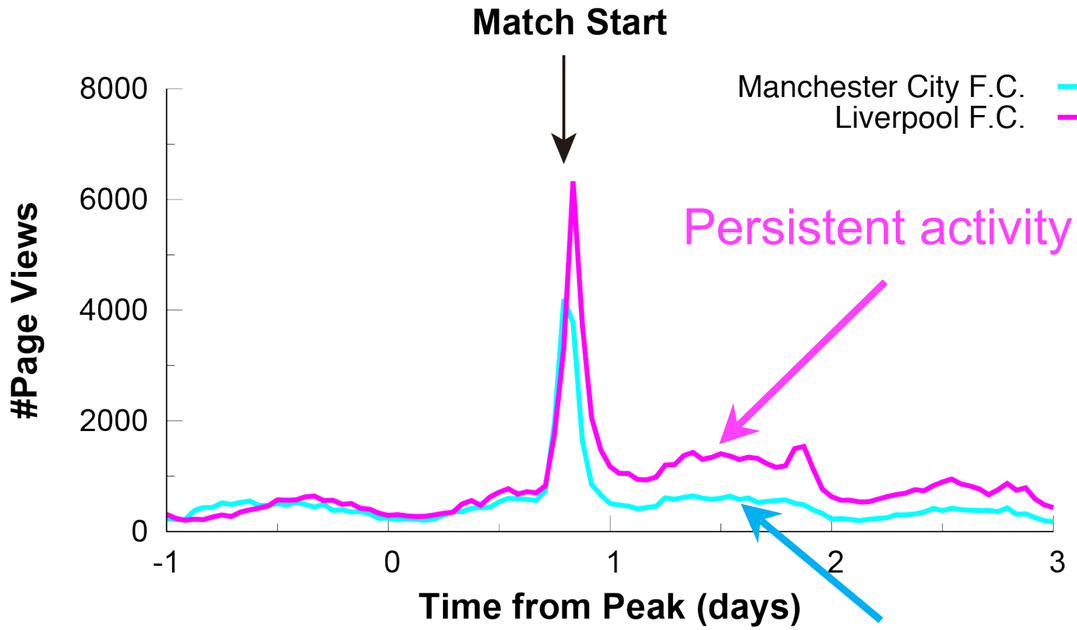
3. Result

The event category (e.g. election) and the anticipation parameters help to improve the prediction.



Event outcome vs Collective attention

Example: 2017-18 UEFA Champions League: Manchester City vs Liverpool FC.



Liverpool FC won the match!

Persistent activity: Excitement??

Rapid decay: Depression??

Table: Parameter distribution

	a_+	τ_+ (hours)
Win	680 (250−1,300)	8.6 (2.5−12)
Draw	770 (200−1,900)	9.4 (1.0−15)
Lose	1,400 (370−4,600)	1.9 (1.0−7.9)

“No one remembers losers” (by Jurgen Klopp)

Conclusion

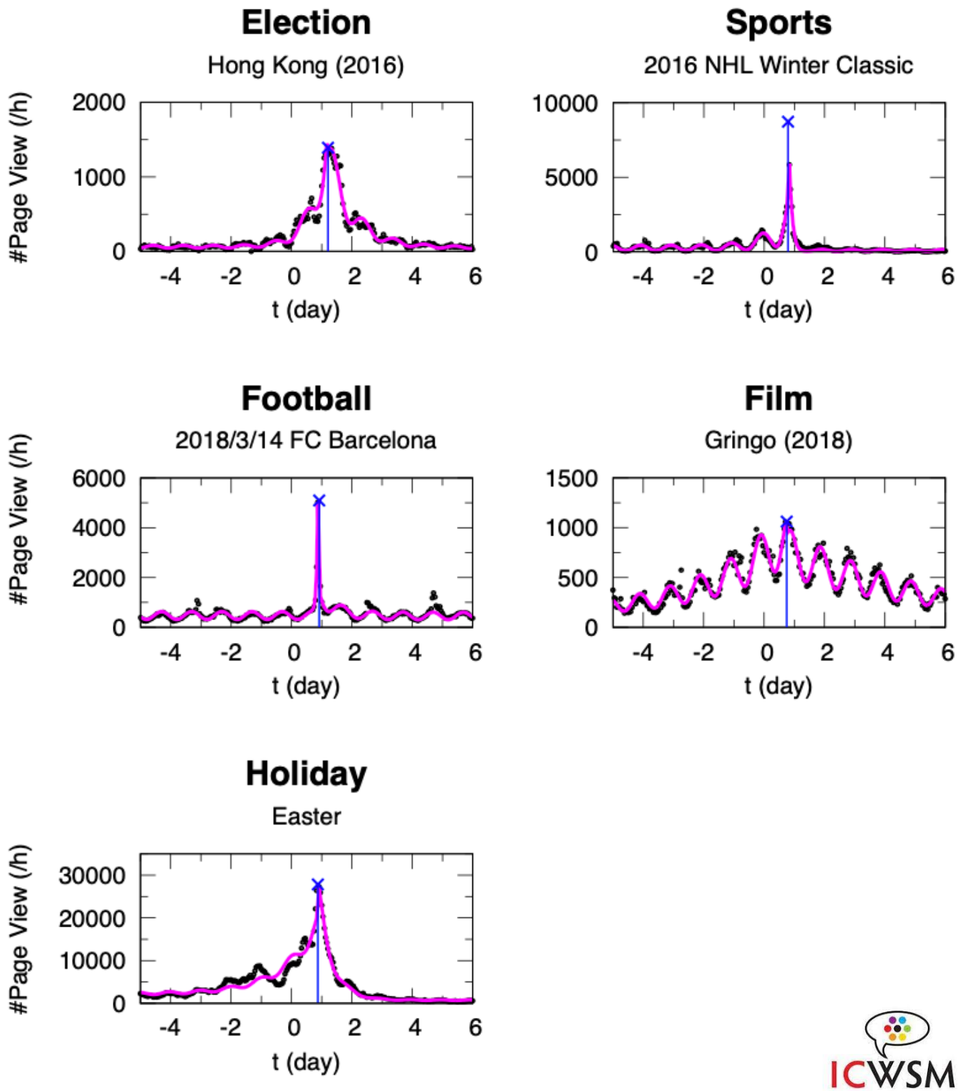
We explore two research questions by developing a simple model for the dynamics of collective attention.

RQ1. What are the essential characteristics of the temporal pattern towards planned events?

1. Exponential anticipation (before the event)
2. Response (after the event), and
3. Circadian rhythms;

RQ2. How is event-related information (e.g., category and outcome) associated with the temporal pattern?

Anticipation + Event information = Response
1. Event category helps to predict the response pattern.
2. Clustering result is associated with the category.
3. Response patterns are associated with the outcome (win, draw, or lose) of the football match.



ICWSM