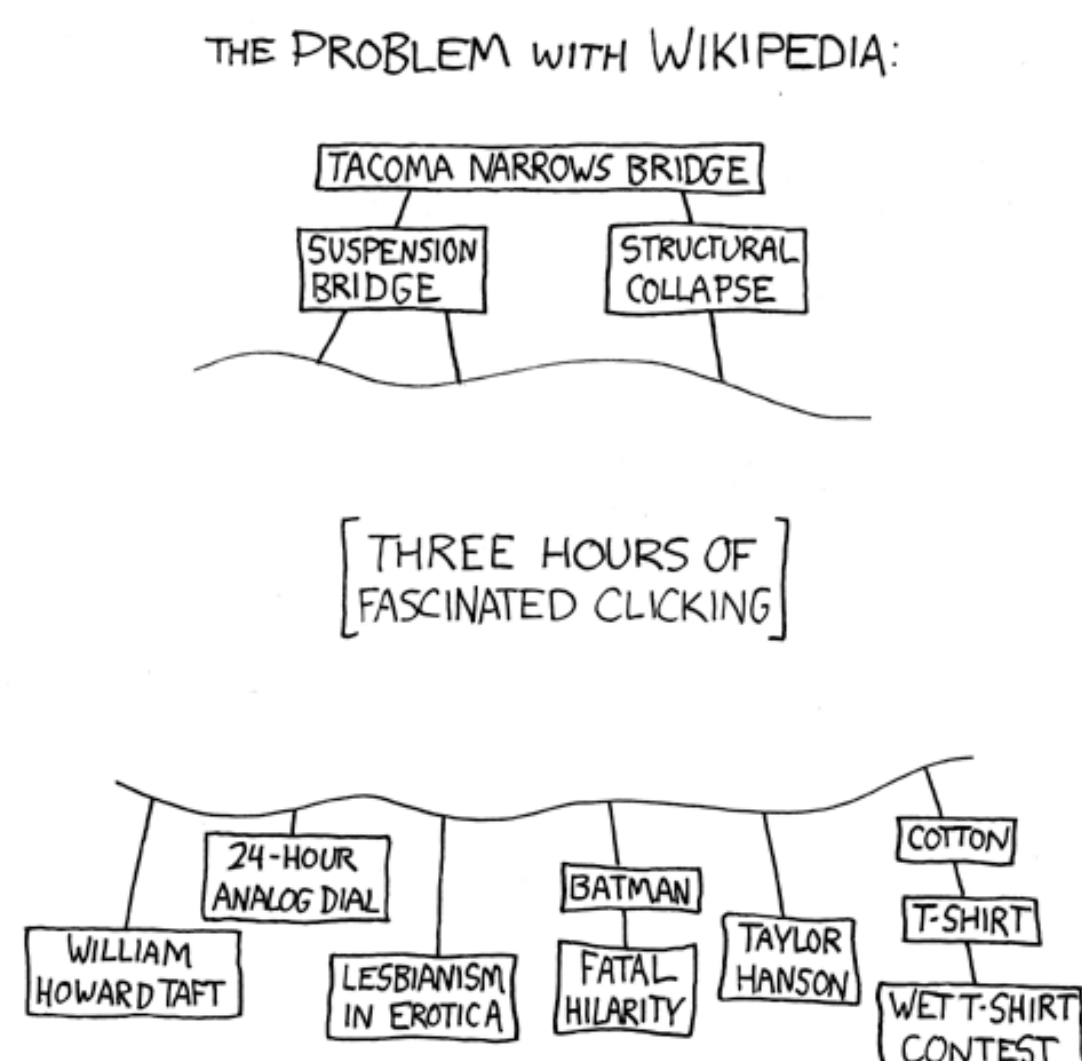


Going Down the Rabbit Hole: Characterizing the Long Tail of Wikipedia Reading Sessions

Tiziano Piccardi, Martin Gerlach, Robert West



Credits xkcd: <https://xkcd.com/214/>

Server logs from English Wikipedia

One month 6.52B load events

Data

User id	Timestamp	Device	Referrer	Article
d6n1t9fgl	2021-04-12 11:29:51	Desktop	bing.com	A
d6n1t9fgl	2021-04-12 11:31:26	Desktop	WP: A	B
d6n1t9fgl	2021-04-12 11:31:33	Desktop	WP: A	C
d6n1t9fgl	2021-04-12 11:36:16	Desktop	WP: C	D
d6n1t9fgl	2021-04-12 11:37:50	Desktop	facebook.com	E
--	--	--	--	--

Summary

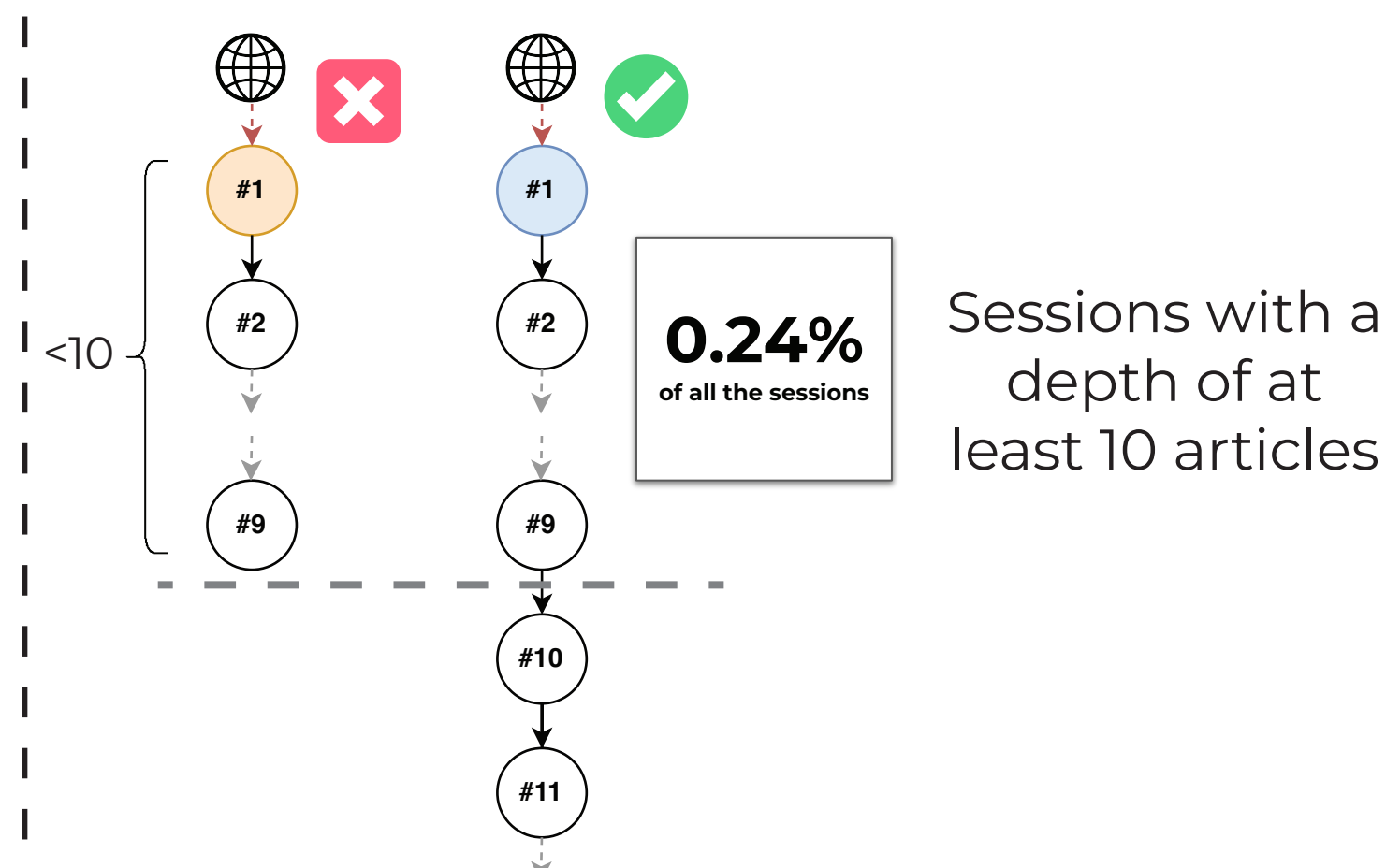
Rabbit hole sessions are **affected** by the **page structure**.

They are more **frequent during night**.

They start **more frequently** from articles about **entertainment**.

On average, they do **not** lead to a **random pages**.

Session aggregation



Frequent starting articles

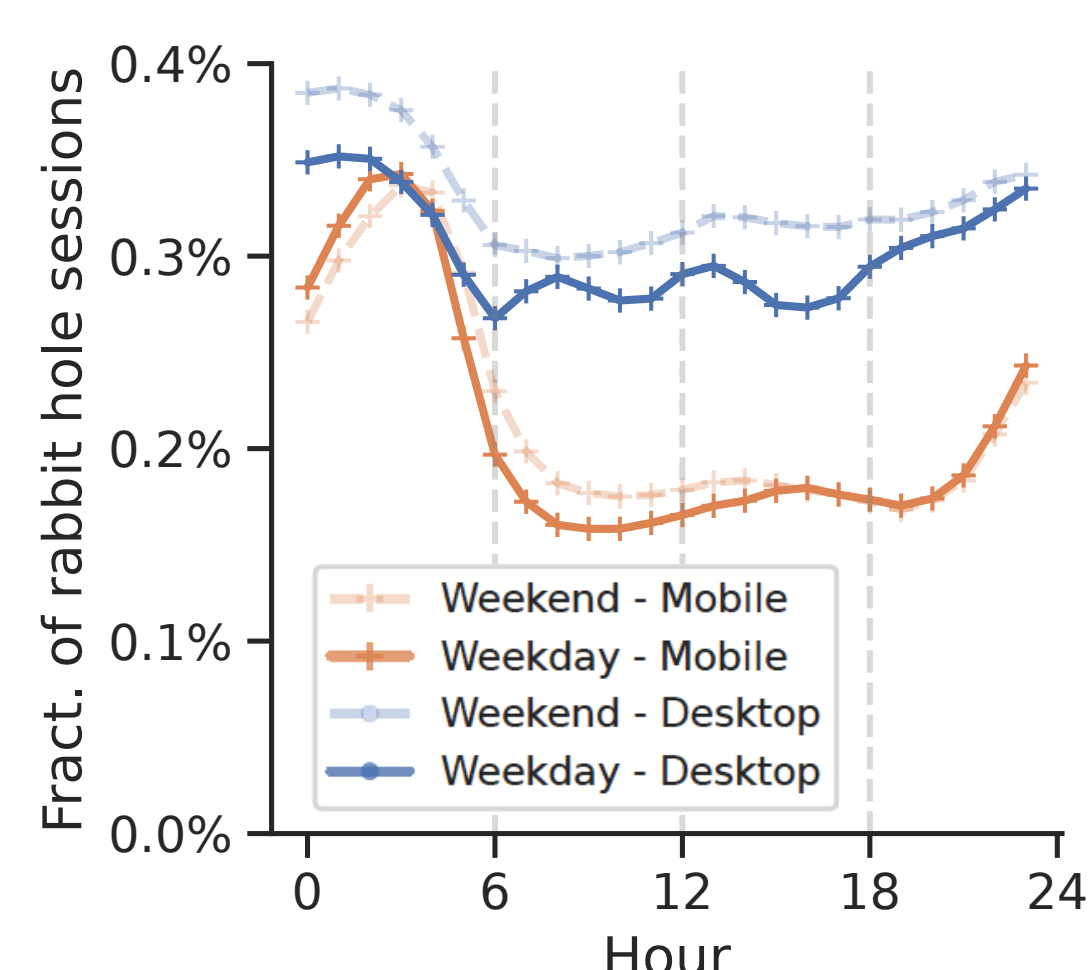
1980 Arkansas gubernatorial election
It is the Law
Alexander, Duke of Schleswig-Holstein-Sonderburg
Solidarity Party of Afghanistan
2006 Florida gubernatorial election

2020 United States presidential election

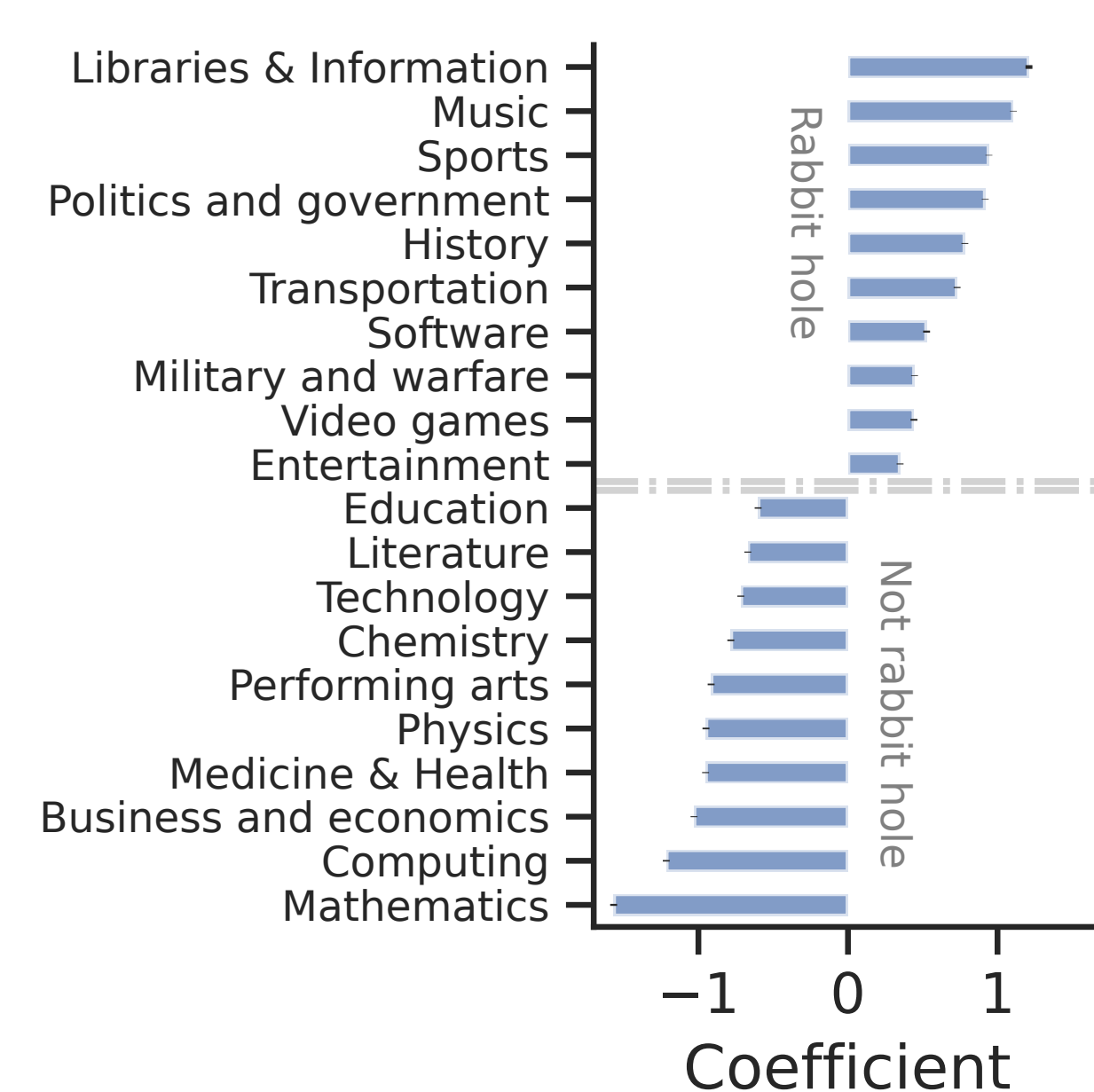
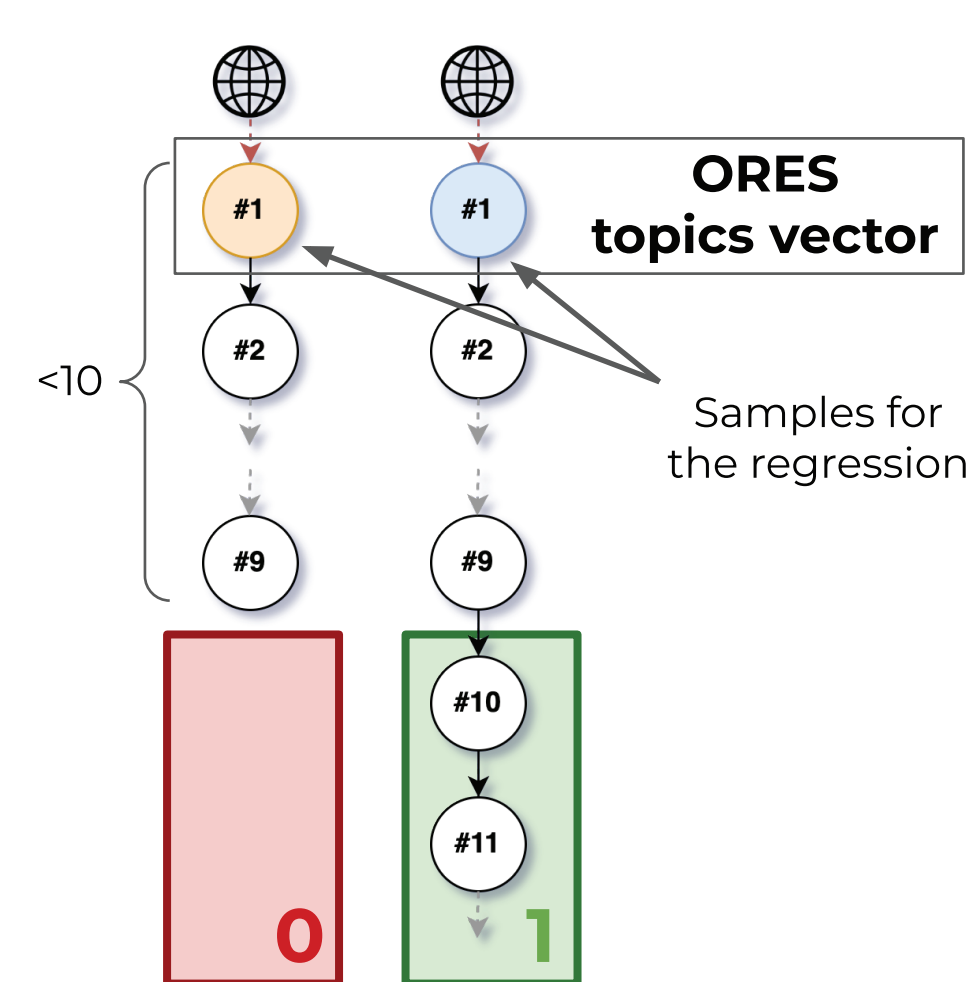
← 2016 November 3, 2020 → 2024 →

Alfred the Great
King of the West Saxons
Reign 23 April 871 – c. 886
Predecessor Æthelred I
King of the Anglo-Saxons
Reign c. 886 – 26 October 899
Successor Edward the Elder

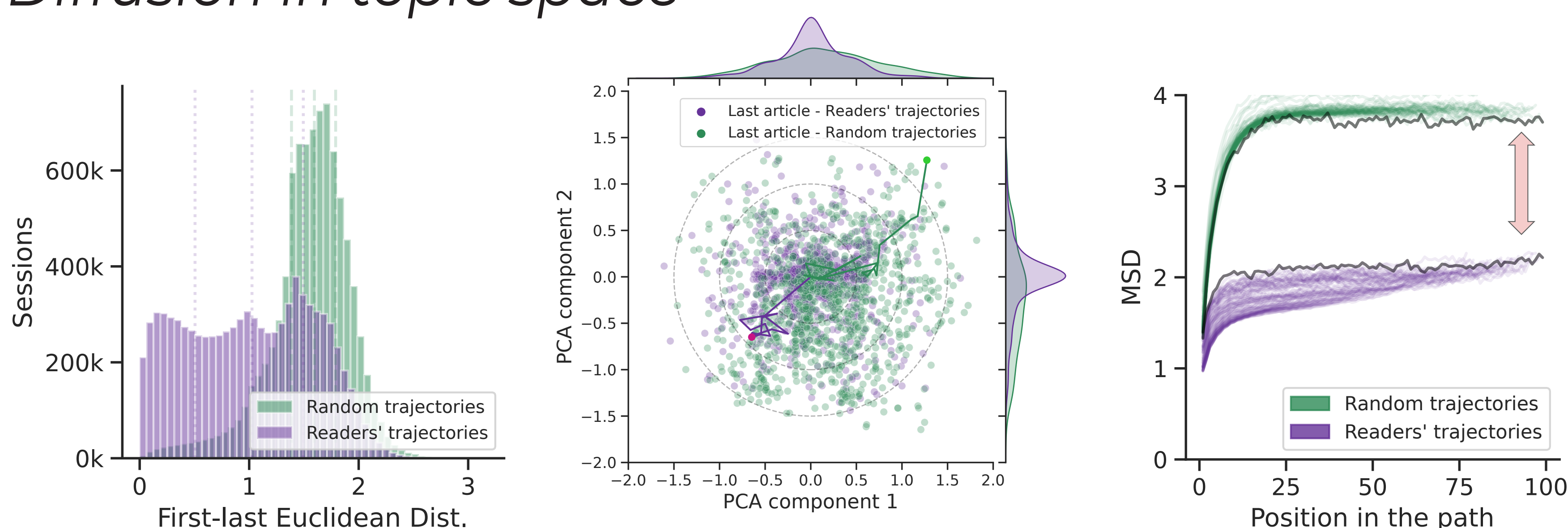
Daily pattern



Frequent topics



Diffusion in topic space



On average, readers do not end up in random pages.

Readers stay in the semantic surrounding of the first page even for long sessions.