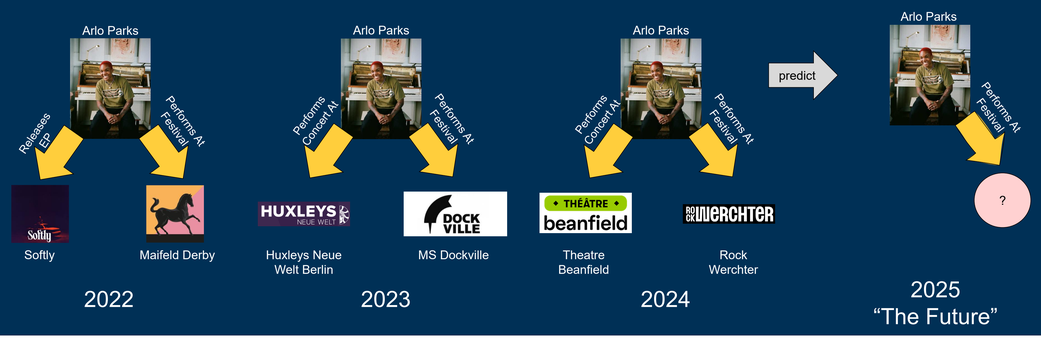
A Temporal Knowledge Graph for Music Festival Lineup Forecasting

UNIVERSITY OF MANNHEIM

Task: forecast festival lineups as link prediction in a temporal knowledge graph (TKG): Which artists will perform a festival next year, and at which festival will an artist perform?

Contributions:

- A TKG of 380 festivals over 55 years, from Setlist.fm and MusicBrainz
- Lineup forecasting as a use-case-driven task with suitable metrics
- Evaluation of six TKG forecasting methods and four zero-shot LLMs



1 Motivation: Why Forecast Lineups?

Fans & organizers

Festival passes are cheapest before the lineup is announced, and fans plan their summer around their favorite bands.

Who will perform at Southside Festival 2027?
(?, performsAt, SouthsideFestival, 2027)

Where will Arlo Parks perform in 2027?
(ArloParks, performsAt, ?, 2027)

Need: future artist–festival links

TKG forecasting research

- Benchmarks (ICEWS, Polecat, GDELT, YAGO, Wikidata) score generic link prediction aggregated over all relations
- Hard to interpret from an application perspective

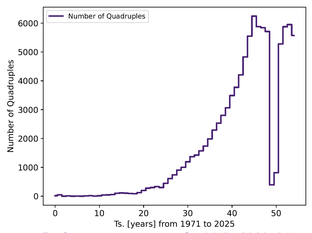
Need: use-case-grounded benchmarks

Can we forecast future lineups from historical data with TKG forecasting?

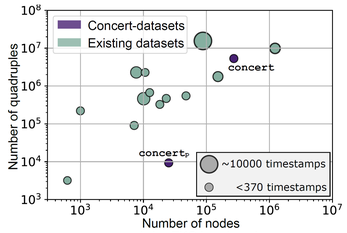
2 Dataset: A TKG of German Festivals

- **Sources:** Setlist.fm (concerts, venues, festivals) + MusicBrainz (artist origin, type, releases, genres, labels), joined via shared artist IDs
- **Scope:** 380 German festivals held ≥5 times with ≥30 artists, 1971–2025, yearly, e.g. *(Arlo Parks, performsAtFestival, Melt!, 2022)*
- **Schema:** 23 relations: performances, releases, artist metadata, labels
- **Two Variants** with identical test queries: `concert` (all relations), $\texttt{concert}_p$ (only *performsAtFestival*)

	Domain	Relation	Range
Performance Information	ARTIST	**performs at festival**	FESTIVAL
	FESTIVAL	happens in venue	VENUE
	ARTIST	performs concert at	VENUE
	VENUE	venue has location	AREA
	AREA	location has country	AREA
Release Information	ARTIST	releases album	RELEASE
	ARTIST	releases EP	RELEASE
	RELEASE	has genre	GENRE-TAG
Meta Data	ARTIST	has begin area	AREA
	ARTIST	has area	AREA
	ARTIST	has type	ARTIST-TYPE
Label Relationships	ARTIST	label founder	LABEL
	ARTIST	recording contract	LABEL
	ARTIST	personal label	LABEL
	ARTIST	artists and repertoire position at	LABEL
	ARTIST	personal publisher	LABEL
	ARTIST	owner	LABEL
	ARTIST	producer position at	LABEL
	ARTIST	creative position at	LABEL
	ARTIST	executive position at	LABEL
	ARTIST	engineer position at	LABEL
	ARTIST	named after label	LABEL
	ARTIST	named after artist	LABEL



Performances per year; Covid dip 2020–21.



`concert` vs. existing TKG benchmarks.

Only 2.9% of 2025 lineup facts recur from 2024 (contrast: YAGO: 92.7%): Re-predicting same lineup as previous year is thus not promising.

3 Task & Evaluation

Single-step forecasting of *performsAtFestival* for 2025, in two directions:

- **Artist prediction** *(?, performsAtFestival, Y, 2025)*
- **Festival prediction** *(X, performsAtFestival, ?, 2025)*

Why time-aware filtered MRR misleads here

Example: A festival with 100 artists; the model ranks a correct artist at every odd position (1, 3, …, 199), a wrong one at every even position.

filled = correct, empty = wrong; correct artists get filtered ranks 1, 2, …, 100

Time-aware filtered MRR
(1 + ½ + … + 1/100) / 100 = **5.2%**

-> Every second prediction correct, yet MRR only 5.2%.

A festival query can have over 200 correct answers.

Additional Metrics:

- **R-Precision (PR):** correct share of the top R_i predictions, R_i = number of true answers
- **Normalized 10-Precision (PN10):** correct among the top 10, divided by min(R_i, 10)

PR = 50/100 = **50%**
PN10 = 5/10 = **50%**

4 Compared Methods

6 TKG forecasting methods

- Rule-based: CountTRuCoLa, TLogic
- GNN + RNN: RE-GCN, CEN
- Path-based: CognTKE
- Recurrency Baseline

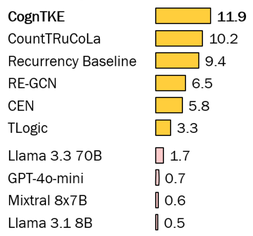
4 zero-shot LLMs, no TKG access*

- Llama 3.1 8B / 3.3 70B, Mixtral 8x7B, GPT-4o-mini
- knowledge cutoff ≤ late 2023
- free-text names are mapped back to TKG nodes

* Question: can broadly accessible, general-purpose LLMs compete with specialized TKG forecasting methods?

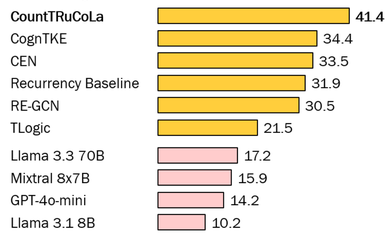
5 Results: PN10 (%)

Predicting artists for a festival



→ hard: too low for standalone use

Predicting festivals for an artist



→ usable as a ranked shortlist

TKG methods, trained on $\texttt{concert}_p$ · zero-shot LLMs, no TKG. · MRR, Hits@k, PR: see paper.

6 What We Learned

Findings on LLM Predictions

- **Popularity bias***:* GPT-4o-mini ranks Billie Eilish top-3 for 47.1% of queries, though she played no German festival in 2025
- **Duplicates***:* Llama 3.1 8B repeats Robin Schulz 23× per list on average
- **Autoregressive effects:** consecutive predictions share properties such as gender or genre

1. Saxon
2. Judas Priest
3. Iron Maiden
4. Scorpions
5. Helloween
6. Blind Guardian
7. Accept
8. U.D.O.
9. Sabaton
10. Powerwolf
11. Rhapsody of Fire
12. Hammerfall
13. Stratovarius
14. Sonata Arctica
15. Dragonforce
16. Alestorm
17. Korpiklaani
18. Eluveitie
19. Ensiferum
20. Finntroll
21. Turisas
22. Sabaton
23. Amon Amarth
24. Children of Bodom
25. In Flame s
26. Dark Tranquillity
27. At the Gates
28. Arch Enemy
29. Nightwish
30. Epica
31. Within Temptation
32. Delain
33. Leaves' Eyes
34. Xandria
35. Lacuna Coil
36. Evanescence
37. Kamelot
38. Symphony X
39. Evergrey
40. Threshold

Llama 3.3 70B for Keep It True Festival 2025: female(-fronted) acts in green; ranks 28–36 are all green.

TKG methods

- CountTRuCoLa can only learn rules of length 1, **no multi-hop dependencies**
  - Still learns plausible rules, e.g.
    performs at Mayfair Ballroom → plays at Rock am Ring
    contract with Spinefarm Records → plays at Wacken Open Air
- TLogic cannot learn **constants** in rules
  - Substantially below the other TKG methods, although competitive on standard benchmarks
- Embedding-based Methods have **out-of-memory errors** on `concert`
  - RE-GCN and CEN underperform Recurrency Baseline
  - CognTKE competitive even when trained only on $\texttt{concert}_p$ -> research into **scalability** worthwhile direction
- **Accuracy grows with artist/festival activity**; for rare entities all methods ≈ baseline

Relations beyond festival performances

- Additional relations (`concert`) change PR/PN10 by at most one point
- RE-GCN, CEN and CognTKE run out of memory
- Slight gains for artist prediction, no consistent pattern for festival prediction
- Enriched data helps in some cases; also adds noise that current models struggle to cope with

Limitations

- Potential coverage gaps in crowdsourced sources (a sample checked against official festival websites showed no major omission)
- Festivals in Germany only; a global TKG would exceed what current models can process
  - Script provided to mine other countries/continents

Future Work

- More scalable TKG models for large real-world event data
- Other realistic queries: partial lineups (remaining performers) & partial schedules (additional bookings)

Paper:

Julia Gastinger, Thilo Dieing, Christian Meilicke, Heiner Stuckenschmidt
University of Mannheim · first.last@uni-mannheim.de · AKBC 2026

Code: github.com/JuliaGast/festival-lineup-tkg · Data: madata.bib.uni-mannheim.de/822 · Paper: arxiv.org/abs/2609.24467

GitHub: