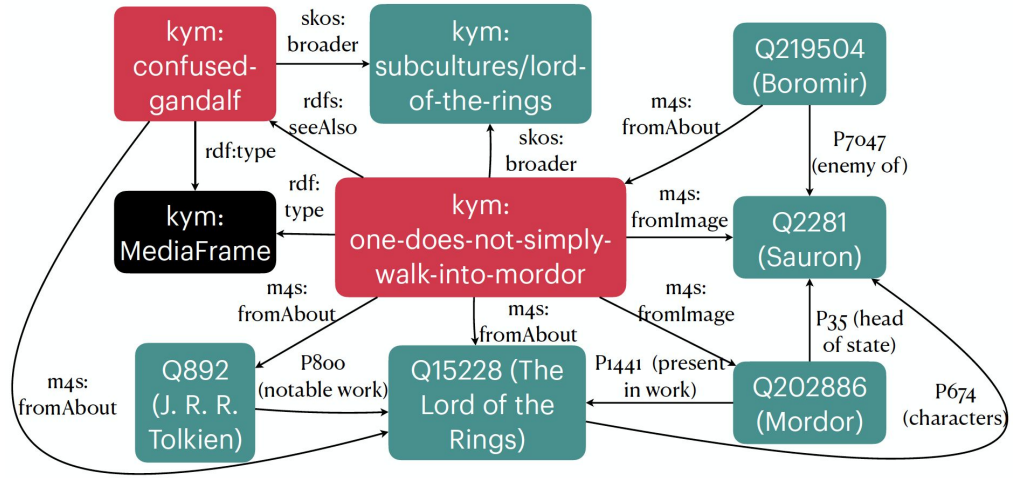


# Lab Program

Time (EST)	Content	Speaker
10:45 - 11:00	Welcome and introduction	Filip
11:00 - 11:20	Internet Memes: <i>knowledge connects culture and creativity</i>	Filip
11:20 - 11:40	Financial transactions: <i>Detecting anomalies in trading</i>	Ke-Thia
11:40 - 12:00	PubGraphs: <i>What should I read next?</i>	Kian & Jay
12:00 - 12:20	Morality in events: <i>From news to timelines and graph maps</i>	Gleb
12:20 - 12:30	Discussion and Closing remarks	Jay

# Refresher! Internet Memes



## a) Frame

Difficulties\_in\_action: factor(s) negatively affecting the ability to achieve objective

## b) Meme

One does not simply *analyse internet memes*



## c) Template

One does not simply walk into Mordor



## d) Origin



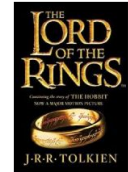
The picture is from **Peter Jackson's** 2021 movie "LOTR: The fellowship of the ring" where **Sean Bean** played "**Boromir**"



## e) Lore



The **book series** "Lord Of The Ring" by **J. R.R. Tolkien** features the character Boromir and the fictional region of **Mordor**



# Run it yourself!

---

<https://github.com/usc-isi-i2/kgtk-aaai2023>



## Welcome

This page contains the notebooks corresponding to the tutorial [KGTK: User-friendly Toolkit for Manipulation of Large Knowledge Graphs](#) given at AAAI'23.

## Installation

To install KGTK, run `pip install kgtk`. Note: there are known issues with Python >3.9, so we suggest using a virtual (conda) environment with Python v3.9.

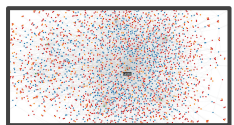
To run the notebooks locally, you can use Jupyter Lab, which is installed with `conda install -c conda-forge jupyterlab`.

If you run into problems, please visit [the KGTK GitHub page](#) for other installation possibilities. If the problems persist, please open an issue on the KGTK GitHub page and we will take a look.

## Use case notebooks

1. [Internet Memes](#) - we show how KGTK can help connect the dots between internet meme sources and external knowledge graphs, like Wikidata. We use KGTK to perform scalable analytics of the resulting graph and execute novel entity-centric and hybrid queries.
2. [Financial transactions](#) - we describe how KGTK can be used analyze financial transaction data. We illustrate how to construct KGTK pipelines with graph transformations, analytics and visualization steps for the financial sector. The KGTK pipelines enable us to highlight trading behaviors, to find potential colluders, and to find inconsistencies through differentiating knowledge graph structures.
3. [Publication graphs \(PubGraphs\)](#) - The recent advent of public large-scale research publications metadata repositories such as OpenAlex (Priem, Piwowar, and Orr 2022) enables us to study innovation at scales that have not been possible before. However, dealing with these large-scale repositories is extremely difficult and requires special toolkits. In this notebook, we describe how KGTK can be used for data filtering, data transformation, knowledge graph extraction, and knowledge graph embedding training of knowledge graphs with scientific publications.
4. [Morality in events](#) - we will demonstrate how our knowledge graph tools are applied to make sense of complex events. Focused on a specific domain (or location) we track the changes in moral foundations (Johnson and Goldwasser 2018) and emotions to understand public perception of these events. The use of KGTK in this project makes it easy to scale up, to generalize to other domains and locations, and to browse and visualize the data. This notebook can be run in [Google Colab](#)

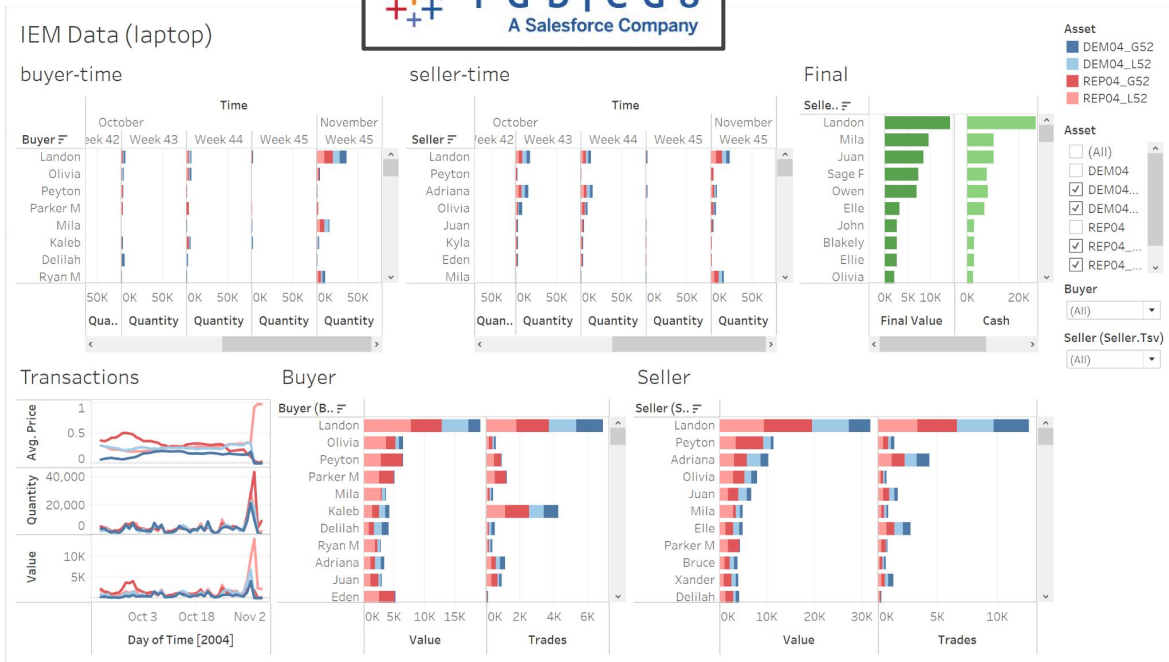
# Refresher! Financial Transactions



transaction  
KG



- analytic: compute balances
- filter: optional
- reusable Tableau dashboards



# Refresher! PubGraphs

Dataset	Nodes	Edges	Rel
OGBL-Cite	3M	30M	1
Freebase	86M	339M	15000
WikiKG90Mv2	91M	601M	1000
PubGraph	432M	15B	51
PG-1M	3M	22M	4
PG-10M	25M	315M	4
PG-Full	184M	2.2B	4

```
titles.head(n=5)
```

	id	node1	label	node2
0	W2138810412-V00_118161	W2138810412	P1476	Content Based Rate Estimation Using Lazy Membe...
1	W1546627792-V00_118162	W1546627792	P1476	Probability Approximations via the Poisson Clu...
2	W2594445009-V00_118163	W2594445009	P1476	A Study on Massive Open Online Courses in High...
3	W2092634194-V00_118164	W2092634194	P1476	Massive Open Online Courses on Health and Medi...
4	W4211094263-V00_118165	W4211094263	P1476	Preconditioning@en

```
papers_idx = embedding_df["node1"].to_list()
titles_idx = dict(titles[["node1", "node2"]].values.tolist())
```

Now, let's find similar papers to "VECTOR VALUED REPRODUCING KERNEL HILBERT SPACES OF INTEGRABLE FUNCTIONS AND MERCER THEOREM."

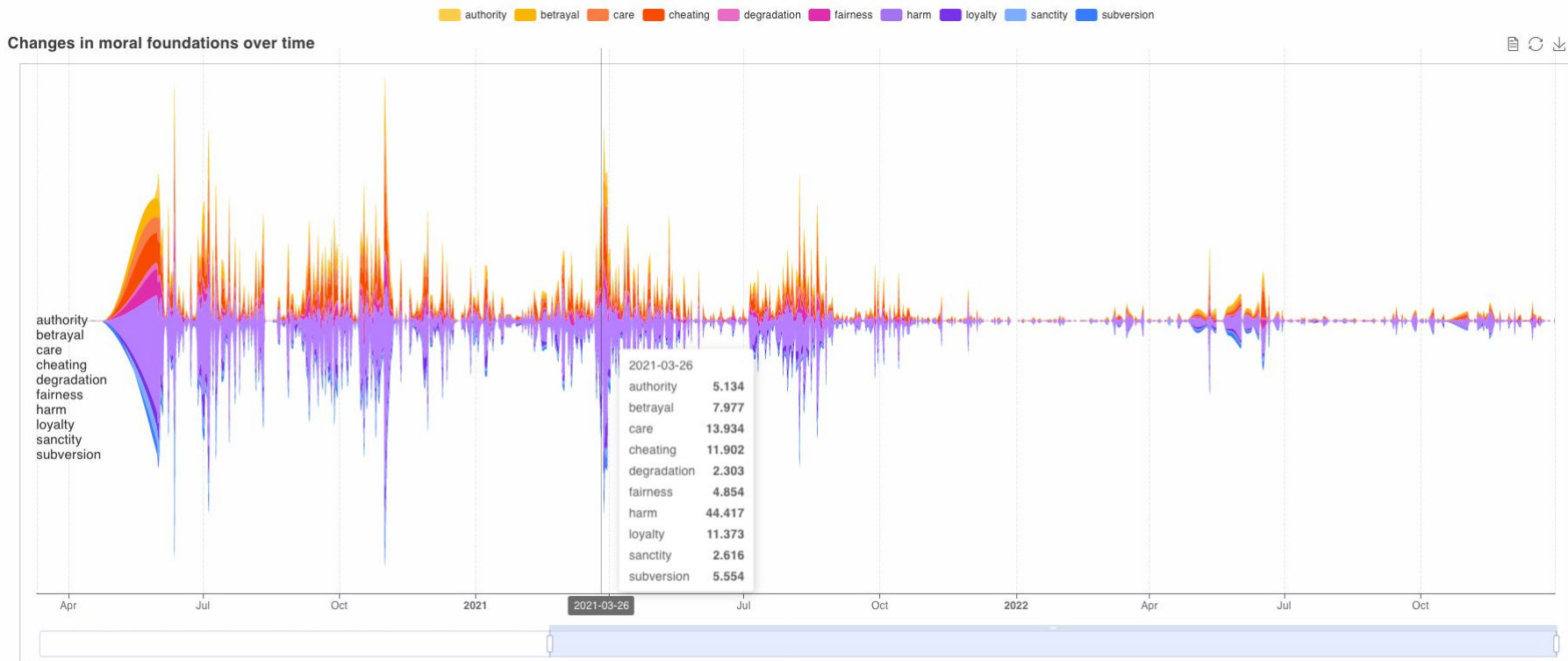
First we need to find the embedding of this paper.

```
candidate = "w2165867509"
candidate_title = titles_idx[candidate]
candidate_embedding = np.array(embedding_df[embedding_df['node1'] == candidate]["node2"].to_numpy()[0]).reshape(1, -1)
```

Then, we can retrieve the closest papers:

```
distances, neighbors = index.search(candidate_embedding, 6)
```

# Refresher! Morality in Events



---

**Other KGTK topics we should have covered?**

**Where should KGTK go next?**

**What did you love (or not love?) about this lab?**

**Other questions?**

**Other KGTK topics we should have covered?**

---



# Where should KGTK go next?

---

**What did you love (or not love?) about this lab?**

---

# Other questions?

---