USING DYNAMOGRAPH: APPLICATION SCENARIOS FOR LARGE-SCALE TEMPORAL GRAPH PROCESSING



Matthias Steinbauer, Gabriele Anderst-Kotsis

Institute of Telecooperation





TALK OUTLINE

- Introduction and Motivation
- Preliminaries on temporal graphs
- DynamoGraph a platform for large-scale temporal graph processing
- How users visit the web
- Real-world Social Networks
- The Global Social Learning Network
- Conclusions



INTRODUCTION / MOTIVATION

- Graphs serve as models for real world structures in many different disciplines
 - Social sciences > Social networks
 - Biology > Protein-Protein Interactions
 - Cartography > Digital road maps
 - Web > The web graph
- Graph and network models are very well studied in mathematics and related disciplines, and computer science
- So why is there need for new research in this area?

REAL WORLD GRAPHS OFTEN GROW TO LARGE SCALES

Exceed memory size of single computer

> require new tactics for visualisation

It is not feasible to process large graphs with **traditional tools** and **algorithms**

THE DIMENSION OF TIME CANNOT BE NEGLECTED

Static views on graphs show often show blurred or too dense data Biological processes are time dependent

Static reachability measures in social networks do not hold for dynamic networks



TEMPORAL GRAPHS

Graph *G* is a pair (*V*, *E*) where *V* denotes the set of vertices and *E* denotes the set of edges between any *v*, $e \in V$

A **temporal graph** *T* can be given as a set of graphs $T = \{G_1, G_2, G_3, ..., G_t\}$ where each $G_x = (V_x, E_x)$ G_x is called a **static snapshot** at **time** x And $G_{tm..tn}$ as a selection of multiple G_x from *T* is a static snapshot for a **timespan**

}

```
id: 39827736,
name: 'Rob Henderson',
description: '',
inEdges: [ {
   weight: 7.3,
   edgeType: 'PHONE',
   source: 39761932,
   target: 39827736, } ],
outEdges: [ {
   weight: 10.0,
   edgeType: 'EMAIL',
   source: 39827736,
   target: 39761932, } ],
```



VERTEX CENTRIC EMBODIMENT WITH TEMPORAL DOCUMENT

```
id: 39827736,
resolution: 'MONTHS',
'1420070400': {
   name: 'Rob Henderson',
   description: '',
   inEdges: [ {
      weight: 3.3,
      edgeType: 'PHONE',
      source: 39761932,
      target: 39827736, } ],
   outEdges: [ {
      weight: 4.0,
      edgeType: 'EMAIL',
      source: 39827736,
      target: 39761932, } ],
},
'1422748800': {
   inEdges: [ {
      weight: 4.0,
      edgeType: 'PHONE',
```

HORIZONTAL SCALABILITY





PREGEL / COMPUTE AGGREGATE BROADCAST



J⊻U

G. Malewicz, M. Austern, A. Bik, J. Dehnert, I. Horn, N. Leiser, and G. Czajkowski. Pregel: A System for Large-Scale Graph Processing. In ACM SIGMOD International Conference on Management of Data, 2010.

DYNAMOGRAPH ARCHITECTURE



HOW USERS VISIT THE WEB

- Center for Complex Networks and Systems Research at Indiana University Bloomington collected the <u>Click Dataset</u>
- **53.5 billion** HTTP headers logged from Sept. 2006 to May 2010
- Available as a ~2.8 TB compressed (~12 TB uncompressed)
- Imported to DynamoGraph in monthly resolution
 - retained only human generated traffic
 - created vertices based on **domain names** (3.4 million after cleansing)

964,243,722

Websites online right now



HOW USERS VISIT THE WEB

- Click temporal graph was successfully loaded into DynamoGraph
 10 compute nodes with 8 threads (graph partitions) each
 On top of an OpenStack cloud
- Currently experiments with distributed PageRank are conducted
 PageRank is computed in sliding windows of 6 to 3 months
 - It is expected that popularity trends in web-sites, ad-networks, and social networks are visible in the data
- First look into the data clearly shows the drop of popularity for the online social network <u>MySpace.com</u>



REAL WORLD SOCIAL NETWORKS

- Users interacting with each other (online or in real-life) form a social network
- Much research from diverse disciplines (sociology, mathematics, computer science, reality mining, …) was already conducted in this area
- Social networks can be modelled as temporal graphs
- Many social interactions are now happening online > easy to track and record
- In collaboration with Ecker (iiWAS 2015) data from Internet Relay Chat (IRC) was logged, annotated and imported to DynamoGraph





full time-span





REAL WORLD SOCIAL NETWORKS

- Data is available as online resource
- Users can filter and load data into DynamoGraph
- Algorithms for automatic layout of the visualisation are available (ForceAtlas2)
- In visualisation it is already clear that reachability measures computed on the full time-span will often not hold on shorter timespans
 - How is information dissemination influenced by that fact?
 - Can we perform cluster analysis on users and identify topics of interest?



THE GLOBAL SOCIAL LEARNING NETWORK



THE GLOBAL SOCIAL LEARNING NETWORK



Ινι

- Experience API (xAPI) used to store learning logs independent of a certain Learning Management System (LMS)
- Scrapers available that convert data from other systems
- An xAPI proxy service is available that can ingest xAPI statements into DynamoGraph
- Setting up experiments with real students

CONCLUSIONS AND FUTURE WORK

- DynamoGraph as a platform for large-scale temporal graph processing has matured enough to be evaluated in scientific scenarios
- Three scenarios were discussed that motivate the need for temporal graph analytics and layout the paths for our future work
- More temporal graph algorithms are to be implemented (reachability, clustering, ...) to provide more interesting metrics to our users



Matthias Steinbauer matthias.steinbauer@jku.at

slides available at <u>http://steinbauer.org/</u>



