In this paper, data refers to the data arising in the context of reasoning challenge to gain insights through traditional database queries and uncertainty of the data. As this big data gets bigger, it becomes a nature, Variety for its heterogeneity, and Veracity for the scale of the data, Velocity for its streaming or dynamic nature. The web communities that includes both the embedded structural information as well as the data generated by users. Web based communities generate such large amount of data that includes both structural changes to the network and updates that are associated with the nodes or the edges of the network. The task of querying and reasoning refers to consumption and management of generated data, and most of the time such task is often performed in real-time as the data arrives.

We briefly discuss some of the important trends and research challenges in supporting querying and reasoning on communities’ data in next section.

2. SOME TRENDS & CHALLENGES

From a technical perspective, we need to explore the right types of infrastructures such as a real-time system with the desired expressiveness and scalability. Although we want to carry out our reasoning/analysis in web communities, we still need to store instances of the social underlying graph and be able to access it with good response rates. So, we need graph databases in order to store and make computations on top of huge graphs. Graph Databases provide the means to store massive graphs in a distributed manner, and perform queries as well as complex computations on them. Similarly to key value stores, they still work with a store and process approach and do not provide much support for temporal analyses.

Graphs are extensively used for representing data, with the result that a number of query languages for graphs have been proposed over the past few decades. Well-established query languages such as relational or XML query language are not suitable for graph structured data since they fall short to provide support for specifying graph traversals. By the way, there have been proposals for graph query languages, but so far except SPARQL no other language has secured broad recognition. However the use of SPARQL language has been mostly restricted to RDF datasets. There are some languages have been proposed recently that build upon SPARQL, such as, Streaming SPARQL [7], EP-SPARQL [8], Continuous SPARQL (C-SPARQL) [9].

Rapidly growing social networks, online communities and other graph datasets require a scalable processing infrastructure. MapReduce, despite its popularity for big data computation, is problematic at supporting iterative graph algorithms. The MapReduce framework provides austere abstractions as well as mature infrastructure to utilize very large computer clusters for computation and data management. However, the limited persuasiveness and state modelling limit the suitability of MapReduce for workloads with complex expressions and complex state.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media.
WWW’16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
DOI: http://dx.doi.org/10.1145/2872518.2890583
fields offers novel and stimulating ideas for research. databases, and real-time processing. The convergence of these community analysis, social networks, social data streams, graph stream querying and reasoning over web communities’ data is stream reasoning and querying jobs.

One more way to manage the high update rates is use random sampling to reduce the size of the data that needs to be processed. One could sample at two different levels in a web community: sample from the network structure itself to reduce the size of the graph that needs to be processed, or sample from the updates to the content. Ahmed et al. [5] present a comprehensive treatment of network sampling, both in static and streaming situations.

Another interesting stream query on social data is a sampling to reduce the size of the data that needs to be processed. One could sample at two different levels in a web community: sample from the network structure itself to reduce the size of the graph that needs to be processed, or sample from the updates to the content. Ahmed et al. [5] present a comprehensive treatment of network sampling, both in static and streaming situations.

3. PAPERS IN THE WI&C’16 WORKSHOP

The papers, selected for this workshop, deal with some of the significant issues in the field. The keynote will be presented by Babak Esfandiarion (Carlton University) on Distributed Wikis and Social Networks: a Good Fit.

In the workshop, four papers are selected for presentation. These papers are:

-Enriching how-to guides by linking actionable phrases is presented by Nikolaos Lagos (Xerox Research Centre Europe), Alexandre Chernov (University of Tübingen), Matthias Gallé (Xerox Research Centre Europe) and Agnes Sandor (Xerox Research Centre Europe). In this paper authors present a method for enriching community-specific procedural knowledge entries that can be found on the Web. They achieve actionable phrase extraction with an F-score of more than 67%, and they provide a higher linking performance than state-of-the-art methods.

-Detection of Multiple Identity Manipulation in Collaborative Projects written by Zaher Yamak (INSA de Rouen), Laurent Vercouter (INSA de Rouen) and Julien Saumier (INSA de Rouen). Authors are interested in detecting fake accounts that try to bypass the Online Social Networks regulations. The presented methodology detects 99% of fake accounts (on a base of 10,000) on English Wikipedia.

- Ontological Networks: Mapping Ontological Knowledge Bases into Graphs proposed by Lucas Navarro (Federal University of Sao Carlos), Estevam Hruschka Junior (Federal University of Sao Carlos) and Ana Paula (IBM Research Brazil). In this paper, authors describe a graph structure called Ontological Network which can be used generically to map Ontological Knowledge Bases (OKB). It is shown that Ontological Network are convenient for implementing graph-mining based algorithms to find new facts and to extend the OKB.

-StarrySky: A Practical System to Track Millions of High-Precision Query Intents by Qi Ye (Sogou Inc.), Feng Wang (Sogou Inc.) and Bo Li (Sogou Inc.). StarrySky is a practical system for identifying and inferring millions of query intents with high precision and acceptable recall. The inference algorithm achieves up to 96% precision and 68% recall on daily search requests.

4. REFERENCES


