

# TelarKG: a Knowledge Graph of Chile’s Constitutional Process

Renzo Angles  
UTalca & IMFD Chile  
Talca, Chile  
rangles@utalca.cl

Vicente Calisto  
IMFD Chile  
Santiago, Chile  
vicente.calisto@imfd.cl

Javiera Díaz  
DCC, UChile & IMFD Chile  
Santiago, Chile  
javiera.pdm@gmail.com

Sebastián Ferrada  
IDIA, UChile & IMFD Chile  
Santiago, Chile  
sebastian.ferrada@uchile.cl

Aidan Hogan  
DCC, UChile & IMFD Chile  
Santiago, Chile  
ahogan@dcc.uchile.cl

Alexander Pinto  
PUC Chile & IMFD Chile  
Santiago, Chile  
aupinto@uc.cl

Juan Reutter  
PUC Chile & IMFD Chile  
Santiago, Chile  
jreutter@ing.puc.cl

Carlos Rojas  
IMFD Chile  
Santiago, Chile  
carlos.rojas@imfd.cl

Henry Rosales-Méndez  
IMFD Chile  
Santiago, Chile  
henry.rosales@imfd.cl

Hernan Sarmiento  
IMFD Chile  
Santiago, Chile  
hernan.sarmiento@imfd.cl

Etienne Toussaint  
IMFD Chile  
Santiago, Chile  
etienne.toussaint@imfd.cl

Domagoj Vrgoč  
PUC Chile & IMFD Chile  
Santiago, Chile  
vrdomagoj@uc.cl

## ABSTRACT

In this paper we present TelarKG, a knowledge graph (KG) that consolidates multiple sources of information regarding the Chilean Constitutional process, particularly about the work of the members of the Constitutional Convention. TelarKG contains a wide range of public data, such as the materials from the commissions and information about the members’ interventions and votes. In addition, the KG has been enriched with online user-generated content including posts from social networks related to the constitutional process, and transcriptions from the YouTube videos of the plenary sessions. Among other features, TelarKG leverages the semantic similarity search capabilities of the MillenniumDB endpoint to compute text similarity among convention members’ Twitter messages. Additionally, we employ Entity Linking techniques to identify references to members of the Convention and Parties within these messages. TelarKG constitutes one of the most comprehensive historical archives on a political event, enabling users to query and navigate through diverse aspects of Chile’s constitutional process.

## CCS CONCEPTS

• **Information systems** → **Graph-based database models**; *Digital libraries and archives*.

## KEYWORDS

TelarKG, Property Graphs, Knowledge Graphs, Historical Archives

### ACM Reference Format:

Renzo Angles, Vicente Calisto, Javiera Díaz, Sebastián Ferrada, Aidan Hogan, Alexander Pinto, Juan Reutter, Carlos Rojas, Henry Rosales-Méndez, Hernan Sarmiento, Etienne Toussaint, and Domagoj Vrgoč. 2024. TelarKG: a Knowledge Graph of Chile’s Constitutional Process. In *7th Joint Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) (GRADES-NDA ’24)*, June 14, 2024, Santiago, AA, Chile. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3661304.3661899>

## 1 INTRODUCTION

The constitutional process in Chile has been one of the most significant socio-political events in the country’s recent history. Starting in October 2019 with a series of protests denouncing bad social conditions across the country [2], this constitutional process – aiming to replace the current constitution instituted during the dictatorship of Augusto Pinochet – involved a series of milestones. These milestones included an initial Referendum in which the public voted overwhelmingly in favor of a new constitution, the election of the members of the Constitutional Convention who would draft the new constitution, and the Constitutional Referendum to determine whether the public agreed to adopt the Political Constitution of the Republic drafted by the Constitutional Convention.

With the goal of enriching the public discourse, *Plataforma Telar* was launched as a data-driven project involving both computer and political scientists. The main goal of the project was to weave together data from multiple sources relating to the Chilean Constitutional Convention, to analyze these data from a socio-political perspective, and to inform the Chilean public of the results.

Plataforma Telar started collecting data in May 2021 before the public elections to form the Constitutional Convention, continuing up to the final referendum in September 2022. The project gathered

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*GRADES-NDA ’24, June 14, 2024, Santiago, AA, Chile*

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0653-0/24/06

<https://doi.org/10.1145/3661304.3661899>

data from numerous public sources, including news from all major national sources, roll call votes, speeches from the convention members, and social network data related to the convention from X (formerly Twitter), Facebook, Instagram and YouTube. Plataforma Telar would then use these data to provide weekly analysis that focused on public opinion, the ideological or political organization of the convention, patterns in voting, and the relationships connecting these facets of the process. The analyses were presented live on national TV in a weekly slot on CNN Chile<sup>1</sup>, enabling the results of Telar to reach a substantial number of people.

Data for Plataforma Telar were collected in tabular format as a data lake hosted on Google Cloud. However, as the number of sources grew to meet the demands of the weekly analyses on CNN Chile, and as the diversity of the resulting data grew, it became more and more difficult to manage the data, with key details of, for example, Constitutional Members being spread across multiple disparate tables. For this reason, we built *TelarKG*: a knowledge graph encompassing all of the data collected by Telar, with nodes representing entities of interest, and edges representing relations between these entities. This entity-centric modelling provides a unified view over all of the data available in Telar about a particular entity, and allowed us to leverage an in-house graph database – *MillenniumDB* [10] – to query for key connections between entities.

In this case study, we describe our experiences on constructing and using *TelarKG* in the context of the Chilean Constitutional Process, and more broadly, in the context of the Political Sciences. Section 2 describes the modeling and construction of the knowledge graph from the base data sources. Section 3 describes how the data are indexed and queried via the *MillenniumDB* graph database. Section 4 describes how we have enriched the knowledge graph – taking into account the rich information contained in text – in order to enable further analytics. Section 5 details how we publish *TelarKG* on the Web – as dumps and a query endpoints – providing a novel open knowledge graph in the Political Sciences domain, which we hope will help stimulate further research. Section 6 concludes.

## 2 MODELING AND CONSTRUCTION

To enable a richer, more comprehensive analysis, we built a knowledge graph (KG) from the raw data of Telar, and publish it as a property graph. In addition, as we describe in Section 5, we have also mapped this KG to the RDF data model. In this section, we describe the modeling and construct of the KG, which we call *TelarKG*.

In Figure 1, we illustrate the general schema of *TelarKG*, where each node label is shown in boxes, and the arrows indicate the existing edge types among these kinds of nodes. Further, in Table 1, we present some statistics showcasing the size of *TelarKG*.

The most central nodes of the graph have the label *ConventionMember*, which model the people that were elected as members of the Constitutional Convention to draft the new constitution. These nodes have attributes like name, occupation, and patrimony. *ConventionMembers* are connected to their *Party*, *District*, etc., fully characterizing their political and demographic stance.

The work of the Convention had two phases. During the first phase the Convention prepared the regulation that would govern

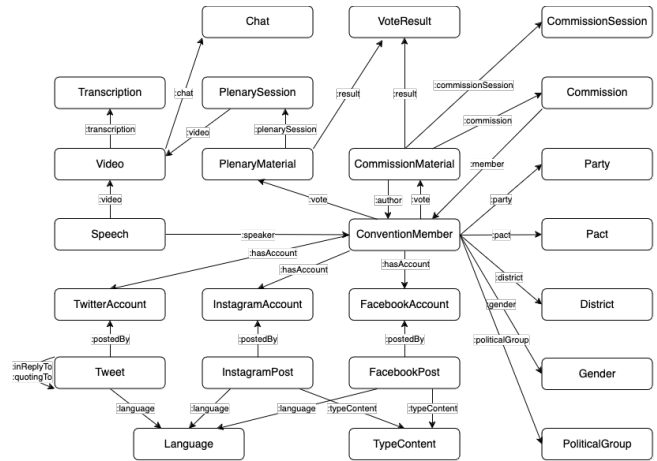


Figure 1: Graph schema of *TelarKG*.

Table 1: Number of nodes and edges of certain types.

Node/edge label	# of elements
ConventionMember	155
CommissionMaterial	8,715
PlenaryMaterial	4,591
Speech	4,099
Video	859
Tweet ( <i>posted by a convention member</i> )	463,605
Tweet ( <i>containing any event-specific keyword</i> )	19,370,342
:vote	874,091
:speaker	4,085

their own functioning, and the second phase consisted on writing the actual constitution draft. To write both the regulation and the constitution draft, the Convention was divided into thematic commissions, which prepared materials that were then voted on in sessions. These commissions are represented by nodes labeled as *Commission*, which are connected to their respective audiences, represented as *CommissionSession* nodes. Similar to commissions, plenary sessions are stored in nodes labeled as *PlenarySession*.

*TelarKG* stores the materials voted in commission sessions and plenary sessions using nodes of type *CommissionMaterial* and *PlenaryMaterial*, respectively. Votes to approve or reject each material are stored as edges with label *vote* connecting the corresponding *ConventionMember* and the *PlenaryMaterial* or *CommissionMaterial*. Each vote edge has an attribute *voteWord* which takes values yes, no, or abstention depending on how the member voted. Each material is connected to a node *VoteResult*, whose name reflects the three possible results of the voting: approved, rejected, or tie.

All of the commission and plenary sessions were live-streamed via YouTube. Metadata of these videos are captured in nodes with label *Video*, which link to the session the video belongs to. During a recorded session, several convention members may speak. These oral interventions are stored in nodes labeled as *Speech*, which include attributes such as the duration of the speech, and the content

<sup>1</sup>See <https://www.cnnchile.com/tag/plataforma-telar/> (Spanish)

```
MATCH (?comm)-[?sp :speaker]->(?member), (?member)-[?pr :party]->(?polGroup)
GROUP BY ?member, ?polGroup
ORDER BY COUNT(DISTINCT ?comm) DESC
RETURN ?member, ?polGroup, COUNT(DISTINCT ?comm)
LIMIT 10
```

**Listing 1: Query to obtain the 10 most active convention members based on the recorded sessions of the commissions.**

of their discourse. The transcriptions of all the videos are also available in nodes with label Transcription, linked to the respective videos. The YouTube live chat of the video is also archived, where each message is represented by a node with label Chat, linked to the corresponding video, and containing the timestamp of the message.

Social media data were also collected. For each convention member, we link to their Instagram, X, and Facebook accounts, where available. These accounts are stored in nodes with labels InstagramAccount, TwitterAccount, and FacebookAccount respectively, along with relevant metadata (e.g., number of followers, profile information, etc.). For each account, we archive all posts in nodes labeled InstagramPost, Tweet, and FacebookPost, where the full text/caption is stored, and metadata as URLs, mentions, likes, etc.

Regarding the content extracted from X, we also store all tweets containing event-specific keywords related to the Constitutional Process (e.g., #convencionconstituyente), obtained through X's search API. Similar to the posts published by members of the Constitutional Convention, we archive this content in nodes labeled as Tweet. However, tweets from people that are not ConventionMember, have no outgoing edges to other types of nodes in TelarKG.

We provide documentation describing every type of node and edge, as well as a full list of all the available properties in TelarKG <sup>2</sup>.

### 3 QUERYING AND ANALYSIS

To enable querying and analytics over TelarKG, we index the KG in MillenniumDB [10]: an open source graph database that supports various graph data models, including property graphs, multilayer graphs [1], and RDF graphs. The database likewise supports two query languages: DGQL, a custom Cypher-like language for querying property graphs, and SPARQL, for querying RDF graphs. In this section, describe the querying of TelarKG as a property graph (further details about the RDF graph are provided in Section 5).

Now, we showcase some use-case queries that can be asked using TelarKG's data. First, we can query the 10 most active convention members within the commission sessions, in terms of speeches. As shown in Listing 1, this result is obtained by querying the number of Speech each ConventionMember is connected to. This information, which is not included in the minutes of the commissions, is acquired in TelarKG through the automatic analysis of video transcripts. Hence, this query can provide additional analysis of the level of engagement and participation of each member in the constitutional process, facilitating a deeper understanding of member's behavior and interactions not apparent in written records.

Another example query is shown in Listing 2. In this query, we ask for the number of convention members from the same political party as a given member who voted differently from them. In this case, we choose Felipe Mena (with node id pers\_mena\_felipe), who belongs to the *Unión Demócrata Independiente (UDI)* party. The

```
MATCH (pers_mena_felipe) <-[?p1 :vote] - (?material :PlenaryMaterial),
      (pers_mena_felipe) - [:party] -> (?y),
      (?x) <-[?p2 :vote] - (?material), (?x) - [:party] -> (?y)
WHERE ?p1.voteWord != ?p2.voteWord
GROUP BY ?material
RETURN ?material, COUNT(?x)
```

**Listing 2: Query to obtain the number of UDI members that had a voting option different to Felipe Mena's.**

results of this query provide an overview that evidences the diversity of opinions among convention members. This is particularly evident when comparing the voting patterns of a specific member with those of others, especially within the same political party.

## 4 ENRICHMENT

The textual content of TelarKG (social media posts, transcripts, etc.) contains rich information that is not reflected in the structure of the KG. Thus, it is not directly possible to analyze the conversation on social media platforms, including questions such as *which convention members are posting on social media about which other convention members, or how diverse is the textual content published by different political parties on social media*. In this section, we describe how we enriched TelarKG to enable a more in-depth analysis of the latent information present in the text nodes of the KG.

### 4.1 Entity Linking

Many social media posts captured by Telar make reference to entities described by TelarKG, but such connections are not directly reflected in the graph structure. In order to make such connections explicit, we applied *Entity Linking (EL)* to the tweets of TelarKG.

As an example, consider the following tweet snippet posted by the convention member Jorge Arancibia, making reference to another convention member, Agustín Squella:

[...] esto originó una propuesta del <sup>Q5660884</sup> Profesor Squella [...]

The goal of EL is to identify mentions (e.g., Profesor Squella) of KG entities in a text, and to assign it the corresponding identifier (e.g., Q5660884 identifies Agustín Squella in Wikidata). This enables the node for the aforementioned tweet to be linked to the related node of TelarKG. Ultimately, enrichment via EL allows us to answer queries about what entities are mentioned in such texts.

We apply two EL systems to enrich TelarKG in this manner: OpenTapioca [3] and DBpedia Spotlight [6]. We choose these systems based on their support for Spanish, their ability to generate (indirectly in the case of DBpedia Spotlight) links to Wikidata, and the availability of an open API. We chose Wikidata as a target since, at the time of enrichment, it described key entities relating to the constitutional process, including all convention members, political parties, etc. TelarKG provides links to Wikidata in such cases, allowing entity links to be propagated from the latter to the former.

Applying Entity Linking on all the tweets in TelarKG, we found 29,311,087 mentions, 117,680 of which are to convention members and 111,840 to parties; we link these latter mentions indirectly to their corresponding nodes, through nodes with label Mention. Though we estimate the precision of DBpedia Spotlight to be 31.5%, and of OpenTapioca to be 43.8% for general mention-entity pairs,

<sup>2</sup><https://telarkg.imdf.cl/documentation>

those pointing to legacy nodes in TelarKG are much higher: the probability of a mention linked to Wikidata being incorrect is much lower when we restrict the range to only those entities in TelarKG, as such entities are named entities, and are all contextually relevant.

The results of EL allow novel analytics over TelarKG. For example, we can now query for those convention members who mention each other the most on Twitter: Cristina Dorador and Elisa Loncón.

## 4.2 Text Embeddings

Social media data offers a valuable resource for exploring how people communicate and express their opinions online. For example, understanding social media discussions can provide insight into public opinion, emerging trends and collective behaviors related to specific issues, especially during socio-political events [9]. Given the diverse sources of online content linked to each convention member, we select X's data to analyze political discourses in this platform. Using Natural Language processing (NLP) techniques, we aim to reveal conversation dynamics and thematic divergences within the discourse of political figures in this constitutional process.

We use the semantic similarity search feature of MillenniumDB, which allows to efficiently compute text similarity within a query. This operation allows us to compute text similarity between the tweets of convention members and quantify thematic divergences across several dimensions, such as temporal or political. We compute vector representations of the posts using a pre-trained model (paraphrase-multilingual-MiniLM-L12-v2) from S-BERT [8].

As an example of the semantic similarity search applied to the tweets of TelarKG, in Listing 3 we show a query that computes the average dissimilarity between a given tweet and all other tweets on a given time frame, grouped by political group. This query enables us to determine how similar the tweets posted by members of a political group were when compared to a specific topic, potentially relevant for online discussions within a defined time range.

## 5 PUBLICATION

After the Constitutional Referendum, in which the Chilean voters rejected the new Constitution, the Telar project began to wind down, and the weekly slot on CNN Chile ended. However, multiple questions can be analyzed for future multidisciplinary research, spanning both Political and Computer Science. The dynamics of the formation of public opinion, the role of digital campaigns in influencing voters' decisions, and the effectiveness of online platforms in engaging and educating the voters are areas open to exploration. Furthermore, analyzing the data structure and usage patterns within TelarKG could provide insights into how information dissemination impacts political processes. We believe that TelarKG contains latent answers to these questions, which can contribute significantly to the understanding of digital influence on political engagement and decision-making processes. By publishing TelarKG as an open KG, we aim to promote a collaborative research environment in which scholars in political science, computer science, and related fields can delve into the complexities of modern democracy.

## 5.1 Data Dumps

We provide full data dumps of TelarKG in the qm (quad model) format of MillenniumDB, as well as in JSON format.<sup>3</sup> Additionally, to allow people from the Semantic Web community to use TelarKG, we also provide an RDF dump that maps all the KG to an RDF graph. For a primer on RDF refer to Manola et al. [5]. To produce this RDF dump, we use a modified version of the mapping proposed in the work of Nguyen et al. [7], following the algorithm formally described by Khayatbashi et al. [4], which maps property graph nodes and edges to RDF nodes. In our work, we only convert edges in TelarKG to RDF nodes if the edge has properties (e.g., the :vote edges have a property voteWord). If an edge does not have any properties, it is mapped to an RDF predicate. We use a custom RDF vocabulary to describe node classes, object properties, and datatype properties, extending well-known vocabularies where appropriate (e.g., we use the IRI `telar:text` to capture the texts of materials, transcriptions, chats, and posts, which extends `sioc:content`).

All dumps are available in <https://telarkg.imfd.cl/dumps/>.

## 5.2 Query Endpoint

The data can be accessed through a public MillenniumDB endpoint at <https://telarkg.imfd.cl/>. The query responses can be exported in CSV format, providing a structured overview of the requested data. Finally, the endpoint provides several query examples that facilitate users to the understanding of basic graph patterns, property paths, and similarity search as applied over TelarKG's data.

## 6 CONCLUDING REMARKS

This paper introduced TelarKG, an enriched KG that integrates multiple sources of information related to the Chilean Constitutional process, offering a detailed overview of the contributions and discussions among members of the Constitutional Convention of 2021–2022. TelarKG allows users to analyze several dimensions of the Constitutional Process, such as the diversity of voting options for each presented material, or the semantic difference on the social media content regarding the posts shared by convention members. All of these elements contribute to a deeper understanding of the Chilean Constitutional Process and, by extension, the Chilean politics. By analyzing the variety of votes and the differences in how members communicate on social media, TelarKG helps us see the full picture of how decisions are made and discussed. This type of analysis provides insights into the constitutional discussions, making it easier for everyone to follow and participate in the process.

For future work, we plan to publish various text embeddings for all the text fields in TelarKG, such as speech transcripts and the discussed materials. These content-embedded representations will allow us to gain deeper insights into the semantic relationships and thematic structures within the dataset, enhancing our ability to perform multi-source analysis related to the constitutional process. Additionally, we plan on further enriching the graph with EL mentions, particularly for the interventions of the convention members, where they usually mentioned their colleagues.

We believe that TelarKG is a first step towards a meaningful modeling of the complexities and dynamics of political discussion

<sup>3</sup>Due to X's Terms of Service, we are not allowed to share the posts' full text contents, so these are not present in the dumps nor in the query endpoint.

```

MATCH (?tw :Tweet {isRetweet:false})-[:postedBy]->(:TwitterAccount)<-[:hasAccount]-(?convMemb :ConventionMember),
      (twp_1494142390663360512)-[:postedBy]->(:TwitterAccount)<-[:hasAccount]-(?author :ConventionMember),
      (?author)-[:politicalGroupConvention]->(?authorPConv),
      (?convMemb)-[:politicalGroupConvention]->(?convMembPConv)
PROJECT_SIMILARITY(?tw, ?dist, "tweets_sbert", twp_1494142390663360512, ANGULAR)
WHERE ?tw.createdAt >= dateTime("2022-02-17T02:51:20") AND ?tw.createdAt <= dateTime("2022-02-17T14:51:20")
GROUP BY ?convMembPConv
RETURN ?convMembPConv, COUNT(?tw), AVG(?dist)

```

**Listing 3: Query to calculate the average dissimilarity between a specific tweet and all other tweets within a specified time interval grouped by political group.**

at all levels. As such, it can be later extended to compile data about parliamentary discussions and voting, where the process in which a bill becomes law can be systematically archived and analyzed, as well as the social discourse surrounding the political discussions.

## ACKNOWLEDGMENTS

This work was partly funded by ANID - Millennium Science Initiative Program - Code ICN17\_002.

## REFERENCES

- [1] Renzo Angles, Aidan Hogan, Ora Lassila, Carlos Rojas, Daniel Schwabe, Pedro A. Szekely, and Domagoj Vrgoč. 2022. Multilayer graphs: a unified data model for graph databases. In *GRADES-NDA '22: Proceedings of the 5th ACM SIGMOD Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA), Philadelphia, Pennsylvania, USA, 12 June 2022*. ACM, 11:1–11:6. <https://doi.org/10.1145/3534540.3534696>
- [2] Octavio Ansaldi and María Pardo-Vergara. 2020. What constitution? On Chile's constitutional awakening. *Law and Critique* 31 (2020), 7–39.
- [3] Antonin Delpeuch. 2020. OpenTapioca: Lightweight Entity Linking for Wikidata. In *Proceedings of the 1st Wikidata Workshop (Wikidata 2020) co-located with 19th International Semantic Web Conference (ISWC 2020), Virtual Conference, November 2-6, 2020 (CEUR Workshop Proceedings, Vol. 2773)*. CEUR-WS.org. <https://ceur-ws.org/Vol-2773/paper-02.pdf>
- [4] Shahrzad Khayatbashi, Sebastián Ferrada, and Olaf Hartig. 2022. Converting property graphs to RDF: a preliminary study of the practical impact of different mappings. In *Proceedings of the 5th ACM SIGMOD Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) (Philadelphia, Pennsylvania) (GRADES-NDA '22)*. Association for Computing Machinery, New York, NY, USA, Article 10, 9 pages. <https://doi.org/10.1145/3534540.3534695>
- [5] Frank Manola, Eric Miller, Brian McBride, et al. 2004. RDF primer. *W3C recommendation* 10, 1-107 (2004), 6.
- [6] Pablo N. Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. DBpedia Spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems, I-SEMANTICS 2011, Graz, Austria, September 7-9, 2011 (ACM International Conference Proceeding Series)*. ACM, 1–8. <https://doi.org/10.1145/2063518.2063519>
- [7] Vinh Nguyen, Hong Yung Yip, Harsh Thakkar, Qingliang Li, Evan Bolton, and Olivier Bodenreider. 2019. Singleton Property Graph: Adding A Semantic Web Abstraction Layer to Graph Databases. *BlockSW/CKG@ ISWC 2599* (2019), 1–13.
- [8] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. <http://arxiv.org/abs/1908.10084>
- [9] Hernan Sarmiento, Felipe Bravo-Marquez, Eduardo Graells-Garrido, and Barbara Poblete. 2022. Identifying and Characterizing New Expressions of Community Framing during Polarization. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 841–851.
- [10] Domagoj Vrgoč, Carlos Rojas, Renzo Angles, Marcelo Arenas, Diego Arroyuelo, Carlos Buil-Aranda, Aidan Hogan, Gonzalo Navarro, Cristian Riveros, and Juan Romero. 2023. MillenniumDB: An Open-Source Graph Database System. *Data Intelligence* 5, 3 (08 2023), 560–610. [https://doi.org/10.1162/dint\\_a\\_00229](https://doi.org/10.1162/dint_a_00229)