Supplementary Materials for **Characterizing the Users, Challenges, and Visualization Needs of Knowledge Graphs in Practice** (Submission ID 1239).

The contents of this document provide supplementary material for the full list of our interview participants' knowledge graph use cases, their feedback on knowledge graph visualization tools from research, and their provided definitions for a knowledge graph (KG).

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the Department of the Air Force under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Air Force.

1 Knowledge Graph Use Cases

As part of our interview study (described in the main text of our paper), participants told us about their uses cases for KGs. A full list of the use cases discussed with us during our interviews is shown in Table 1.

Use Case	**Description**
Path Discovery	Finding new cyber threat pathways in a cyber KG [HKSR*20] connecting computer systems and known exploits
	Finding new treatment pathways in a KG connecting diseases and possible treatments (e.g., for drug discovery)
	Finding new materials synthesis pathways in a KG connecting different chemical compositions
	Identifying user workflows in a KG connecting user actions in an enterprise network
	Detecting communities, clusters, paths, and semantic proximity in the KG
Data cataloging	Creating a query-able knowledge base for data scientists and developers to quickly find data they need
	Standardizing and de-duplicating terminology and data usage in a knowledge graph
	Allocating resources more efficiently to critical IT infrastructure
	Virtualizing data to improve data retrieval and reduce technical challenges and errors
	Modeling the organization's business logic, e.g. "the organization has X branch, which has Y types of employees."
	Modeling facilities, their security systems, lights, fire suppression systems, etc. to connect to physical floor layouts
	Managing global web content, e.g., showing information about a movie based on the website visitor's country
AI/ML	Identifying key individuals (important nodes) in an enterprise or social network KG
	Explaining why an anomaly was predicted from a model using data that is also connected to the KG
	Predicting stock prices for publicly traded companies in a KG connecting companies, industries, and supply chains
	Using NLP to process text-based data sources to identify entities and connections, and for deduplication
	Using graph neural networks (GNNs) for node and link prediction, regression, and classification)
	Using data integrated from a KG to improve the accuracy and predictive power of ML models
	Training a model to detect author profiles and authoring changes in social media networks

Table 1: A full description of our participants' KG use cases. KGs are used for pathway discovery, data cataloging, as well as training, understanding, and improving AI/ML models.

2 Knowledge Graph Metrics

Table 2 shows an extended participants table that highlights the metrics about the KGs that the participants work with. The first 5 columns match Table 1 from the paper, and the last 5 columns display metrics about the KGs.

PID	Education	Job Title	Company Domain	KG Persona(s)	Est. Number of Nodes	Est. Number of Edges	Est. Number of Properties	Regularly Updates KG	Public / Non-public Data Sources
01	MS	Research Scientist	FFRDC	Builder, Analyst	100 million	13 billion	10	**Yes**	Public
02	MS	Research Scientist	FFRDC	Builder, Analyst	10 thousand	—	—	No	Non-public
03	PhD	Research Scientist	FFRDC	Analyst	10 thousand	—	—	No	Non-public
04	MS	Research Scientist	FFRDC	Analyst	10 thousand	—	—	No	Non-public
05	MS	Research Scientist	FFRDC	Builder, Analyst	100 million	13 billion	10	**Yes**	Public and non-public
06	MS	Research Scientist	FFRDC	Analyst	100 million	13 billion	10	**Yes**	Public
07	MS	Research Scientist	FFRDC	Analyst	500 thousand	3 million	10	No	Public
08	MS	Software Developer	FFRDC	Builder, Analyst	500 thousand	3 million	10	No	Public
09	MS	Research Scientist	FFRDC	Builder, Analyst	5 thousand	2 million	30	**Yes**	Public
10	PhD	Research Scientist	FFRDC	Builder, Analyst	billions	—	—	**Yes**	Public
11	MS	Data Analyst	Enterprise (Finance)	Builder, Analyst	5 thousand	8 thousand	3	No	Public
12	PhD	Director	Enterprise (Health)	Builder, Analyst	50 million	2 billion	200	**Yes**	Public and non-public
13	BS	Industry Analyst	Enterprise (Consulting)	Analyst, Consumer	millions	—	—	No	Public
14	MS	PhD Student	Academia	Builder, Analyst	100 million	13 billion	—	**Yes**	Public
15	PhD	Data Scientist	Enterprise (Health)	Analyst	100 million	13 billion	—	**Yes**	Public
16	PhD	Comp. Biologist	Enterprise (Health)	Builder, Analyst	50 million	2 billion	200	**Yes**	Public and non-public
17	PhD	Principal Scientist	Enterprise (Tech)	Builder, Analyst	billions	—	—	**Yes**	Public
18	MBA	Digital Lead	Enterprise (Health)	Consumer	—	—	10 thousand	**Yes**	Non-public
19	MS	PhD Student	Academia	Builder	millions	—	—	**Yes**	Public

Table 2: Extended participant demographics for our interview study, described in Section 3 of the paper. From left to right: the participant's ID; job title; the organization they work in (FFRDC stands for Federally Funded Research and Development Center); their primary persona(s) as KG users (further explained in Section 4 of the paper); estimated number of nodes in their KG; estimated number of edges; estimated number of properties; whether they regularly update their KG; and whether their KG was created from public and/or non-public data sources. — lines indicated the participants did not provide or know the information.

3 Feedback on Knowledge Graph Tools in Research

At the end of the interview, we walked through three published examples of KG visual interfaces to elicit our participants' feedback on the interface's designs and capabilities.

First, a coordinated multi-scaled visualization [HRGK*21] was shown. The authors' approach includes three separate views for a biomedical knowledge graph: a KG global view which clusters the graph by ontological category using a hierarchical circle-packing layout, a KG local view that displays a flow graph from a user's query search results, and a KG drilled down view that provides text and document evidence for link suggestions. Domain experts who reviewed their tool had positive reactions and appreciated that the tool allowed them to explore the graph data and even formulate queries.

Second, a visualization dashboard (*VisKonnect*) [LAG*21] of EventKG [GD18] was shown. There are four views in VisKonnect: a timeline showing life events of queried historical figures, a map showing a geographic view of those events, a node-link diagram as a subgraph of shared events, an embedded Wikipedia article for highlighted events, and an NLP-based panel that lets users query templated questions to the KG without using any query language.

Third, screenshots from Neo4J Bloom (`https://neo4j.com/product/bloom/`) were shown. Screenshots included node-link diagram representations of open source (Neo4J) graph databases. On these interfaces, different panels for filtering the graph were included, as well as control panels for querying the graph.

As our interview participants lacked the complete context and use cases for these three tools, and have different needs than those the tools are designed for, we discussed what *could* be integrated or improved upon for their own future KG visualization use cases. We discuss their feedback below.

For the multi-scale tool [HRGK*21], we found that our participants' reactions were mixed depending on their domains: (4/19) reacted positively to the tool, while (5/19) were unsure of the organization for the 'global' versus 'local' views and were not sure whether it would be helpful in their workflow. (2/19) did not think the system would work well for their domain-specific use case, as they use a highly connected graph that "*would not render well, even in a zoomed-in localized view.*" P12 told us that, in his experience, end users get lost trying to 'zoom in and out' of a KG:

> We did something similar to this but not as sophisticated. Once you start summarizing things, if people want to drill down, you have to start expanding things, then people get lost. There's a lot of zooming in and out, for example if you click on something then go to a lower level, people get lost as to which level they're at... It was just a lot of work for them to do the exploration of what they wanted to see. In the end doing this in a table was just much easier, since they are more used to it. -P12

Participants generally responded positively to VisKonnect [LAG*21], although some remarked that the dashboard visualizations seemed too narrow, as they only showed relationships between two or three entities at a time. In these cases, participants preferred a more high-level, global visualization for viewing the KG. P5 told us that cartograms (as seen in the dashboard) are well-received by end users:

> I know a lot of stakeholders like seeing knowledge graph data as a map... Even when it's not that useful, it's easy eye candy. -P5

When shown screenshots of Neo4j Bloom's node-link diagrams, participants were quick to point out the hairball-like structure in the graphs. We received similar feedback for these visualizations as we did when discussing NLDs in general (found in the main body of our text). Some of our participants had interacted with Neo4JBloom's filtering and querying control panels, albeit with confusion and difficulty. The overall feedback on these tools highlight the need for additional 'straightforward' KG visualization designs (fitted to its unique use case), as well as example-based filtering capabilities in interactive KG tools.

4 Knowledge Graph Definitions

Finally, we include the full list of knowledge graph definitions provided by our interview participants in Table 3. Varying definitions for knowledge graphs have been proposed in the past for KG-related research [EW16] with the goal of creating a shared language to drive research forward. Our goal with this question in our demographics survey was to establish a similar shared language for KGs within the visualization community, as well as to identify how practitioners think about knowledge graphs.

Overall, we find that the practitioners we interviewed define knowledge graphs in terms of: (1) its ability to store different types of nodes (or entities), and different types of semantic relationships (or edges) between those nodes; (2) its ability to help both humans and machines understand what the data "*actually means*" – e.g., in the context of the data domain. These sentiments are close to the proposed definition of knowledge graphs provided by Paulheim [Pau17]: "*A knowledge graph (i) mainly describes real world entities and their interrelations, organized in a graph, (ii) defines possible classes and relations of entities in a schema, (iii) allows for potentially interrelating arbitrary entities with each other and (iv) covers various topical domains.*"

References

[EW16] Ehrlinger L., Wöss W.: Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS) 48*, 1-4 (2016), 2.

[GD18] Gottschalk S., Demidova E.: Eventkg: A multilingual event-centric temporal knowledge graph. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15* (2018), Springer, pp. 272–287.

[HKSR*20] Hemberg E., Kelly J., Shlapentokh-Rothman M., Reinstadler B., Xu K., Rutar N., O'Reilly U.-M.: Linking threat tactics, techniques, and patterns with defensive weaknesses, vulnerabilities and affected platform configurations for cyber hunting. *arXiv preprint arXiv:2010.00533* (2020).

Knowledge Graph Definitions & Criteria
A graphical model that represents things or entities and their relationships with each other. It is a graph in the sense that it has vertices and edges. The vertices are the things and the edges are their relationships. This allows for mathematical analysis.
A directed or undirected graph that shows the relation between entities. Entities can be linked to entities of the same type or different types with different types of relationships. These relationships are user defined.
Consists of nodes and edges, where nodes often represent entities (e.g., objects, subjects, etc.) and edges represent the relationships between the nodes. For example nodes could be movie titles (e.g., Ant Man, Iron Man) and production companies (e.g., Marvel Studios) and an edge connecting these two types of nodes could be "produced-by". Knowledge graphs can be used to fuse key elements of various data sources to create context that is useful for various information retrieval tasks like question and answering.
Helps organize information in a way that can be useful for both humans and machines. Allows for complex relationships to be visualized/processed on. It should be searchable in some sense based on the encoding of it.
A graph that encodes facts and relationships in a structured manner.
Connecting different data sources in a way that adds more meaning than using each data source separately.
An explicit collection of entities and their attributes and relationships.
A set of {entity - link - entity} triplets that represents a set of information about some domain. Knowledge graphs tend to be large, although they do not have to be, and often their scale hides complexities in the knowledge that they encode.
A knowledge and data structure for synthesizing relational patterns and dependencies of concepts or data samples.
A heterogeneous graph that represents real world entities. It shows the relationship and connectivity among different entities and types of entities.
A data structure that represents data about the relationships between entities (drugs, genes, diseases, ...) in a way that allows to analyze these relationships or allows to help contextualize experimental results.
Maps the relationships between objects (data) and provides information that helps humans and machines understand what the data actually means. A knowledge graph is data plus metadata (or semantic information) linked in a graph.
Shows the relationship between entities. More like if you were to put twitter data in a graph database versus looking at a network graph or a purely mathematical graph. So purely that a bunch of nodes are connect does not make a knowledge graph until you give the connections some sort of context or meaning.
Any semantic graph (a graph with labeled edges) and unique nodes is a knowledge graph.
A collection of nodes and edges where the nodes correspond to entities and the edges correspond to relationships between entities. I believe when people refer to knowledge graphs, they refer more to the types of knowledge that are stored, which typically is a set of domain-specific facts that are linked in some way, representing valuable information.

Table 3: The full list of knowledge graph definitions provided by our interview participants (note that 4 participants chose not to provide a definition in their survey response). Most importantly, a KG preserves the semantics of human knowledge by mapping different types of relationships (edges) between different types of entities (nodes). This ability makes KGs interpretable by both humans and machines.

[HRGK*21] Husain F., Romero-Gómez R., Kuang E., Segura D., Carolli A., Liu L. C., Cheung M., Paris Y.: A multi-scale visual analytics approach for exploring biomedical knowledge. In *2021 IEEE Workshop on Visual Analytics in Healthcare (VAHC)* (2021), IEEE, pp. 30–35.

[LAG*21] Latif S., Agarwal S., Gottschalk S., Chrosch C., Feit F., Jahn J., Braun T., Tchenko Y. C., Demidova E., Beck F.: Visually connecting historical figures through event knowledge graphs. In *2021 IEEE Visualization Conference (VIS)* (2021), IEEE, pp. 156–160.

[Pau17] Paulheim H.: Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web 8*, 3 (2017), 489–508.