# Characterizing Complex Network Properties of Knowledge Graphs

Anderson Rossanez[1][a], Ricardo da Silva Torres[2,3][b] and Julio Cesar dos Reis[1][c]

[1] *Institute of Computing, University of Campinas, Campinas, SP, Brazil*

[2] *Agricultural Biosystems Engineering and Wageningen Data Competence Center, Wageningen University & Research, Wageningen, The Netherlands*

[3] *Department of ICT and Natural Sciences, Norwegian University of Science and Technology, Ålesund, Norway*

Keywords: Knowledge Graphs, Complex Networks, Centrality Measurements.

Abstract: Knowledge Graphs have been established as one the most relevant representations to encode *knowledge*, with relevant applications in the public and private sectors. One common research direction concerning the analysis of created knowledge graphs relies on the assumption that their intrinsic properties and structure are similar to what is observed in complex networks. However, studies concerning identifying typical complex network structures in knowledge graphs are lacking in the literature. This paper bridges this gap by analyzing commonly and recently used knowledge graphs in the semantic web field, seeking to demonstrate their complex network properties. Evaluation involving DBpedia and Wikidata data confirms the occurrence of intrinsic complex network structures in their respective knowledge graphs.

## 1 INTRODUCTION

Knowledge Graphs (KGs) (Ehrlinger and Wöß, 2016) are computational tools that model knowledge through the interrelations of real-world entities in facts using a graph structure. Many large-scale KGs are made freely available, such as DBpedia (Auer et al., 2007) and Wikidata (Erxleben et al., 2014), while others are maintained by companies (Noy et al., 2019), such as Google,[1] for instance. The importance of KGs as a means of knowledge representation has been increasing both in academia and industry. Such importance is proven by the ever-growing amount of applications that take advantage of them in several domains (Zou, 2020; Ji et al., 2022).

Due to their graph-based representation, several studies have considered the computation of complex network measurements as a key methodological procedure in different analysis tasks involving KGs (Dörpinghaus et al., 2022). More specifically, studies concerning the use of centrality measurements computed over KGs have been conducted to determine the relevance of concepts represented by their nodes (Park et al., 2019) or verify how such relevance changes over time (Rossanez et al., 2020). Other examples involve determining features over KGs to support their embedding for machine learning usage (Sadeghi et al., 2021), such as, for instance, in predictive models (Tilly and Livan, 2021).

Real-life complex networks, represented by large graphs, present characteristics that distinguish them from random graphs. Several models of complex networks are available, and sets of observable characteristics describe them. No studies in the literature compare the characteristics of complex networks with those from KGs, which would be valuable to ensure the validity of methodologies such as those that employ complex network measurements in KG-based analyses.

In this study, we take a step forward to bridge this gap. Our objective is to demonstrate that KGs present characteristics of real-life complex network models. By observing such characteristics, we can safely use complex network measurements on KG-based analyses. This article assesses complex network properties based on widely known real-world KGs. To the best of our knowledge, no studies in the literature performed such an evaluation.

This article addresses the following research questions:

[a] https://orcid.org/0000-0001-7103-4281

[b] https://orcid.org/0000-0001-9772-263X

[c] https://orcid.org/0000-0002-9545-2098

[1] https://blog.google/products/search/introducing-knowledge-graph-things-not/ (As of Aug. 2023).



Rossanez, A., Torres, R. and Reis, J.
Characterizing Complex Network Properties of Knowledge Graphs.
DOI: 10.5220/0012257300003598
In *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2023) - Volume 2: KEOD*, pages 119-128
ISBN: 978-989-758-671-2; ISSN: 2184-3228

**RQ1.** How can we assess complex network characteristics on KGs?

**RQ2.** Would widely used KGs present complex network characteristics?

**RQ3.** Is it sound to apply complex network measurements on KGs?

In this study, we consider two complex network models, namely *scale-free* and *small-world* networks. We consider the characteristics of such networks in the scope of the formal definition of KGs, and their intrinsic properties. We conduct an experimental procedure involving real-world KGs, namely *DBpedia* and *Wikidata*. We observed the characteristics of our targeted complex network models in sub-KGs representing entries of such datasets previously used in the literature for centrality measurement extraction. We also observed the same characteristics on instances of a Temporal Knowledge Graph (TKG) to verify if they are also observable over time. The results show the availability of complex network characteristics on KGs.

In short, the contributions of this investigation are twofold:

- We found, for the first time, complex network properties in knowledge graphs;
- We present and discuss an analysis showing that complex network characteristics are observed in widely used KGs, which validate ongoing initiatives concerning using complex network measurements for KG-based analyses.

The remainder of this article is organized as follows: Section 2 introduces underlying concepts in this study. Section 3 presents the related work. Our proposal to observe complex network characteristics in KGs is detailed in section 4. Section 5 describes our experimental evaluation and presents the achieved results, which are discussed in Section 6. Section 7 summarizes our findings and points out directions for future research.

# 2 BACKGROUND

We provide an overview of concepts that are relevant to our formulations: *Knowledge Graphs* (*cf.* Subsection 2.1) and *Complex Networks* (*cf.* Subsection 2.2).

## 2.1 Knowledge Graphs

Knowledge Graphs (KGs) are computational tools applied to represent knowledge regarding entities and their relationships (Paulheim, 2017). KGs have recently been exploited by both the industry and the academia in scenarios that require the representation of large-scale, diverse, and dynamic collections of data (Hogan et al., 2021).

A KG defines the interrelations of entities in facts using (subject, predicate, object) triples. Most KGs use the *Resource Description Framework* (RDF) (Candan et al., 2001) representation. They are composed of a finite number of RDF triples (Faerber et al., 2017), in which each of its constituents is represented by Uniform Resource Identifiers (URIs), literals (which commonly describe the meaning of a URI in natural language), or even blank nodes. A triple may be graphically represented by a vertex (predicate) connecting two edges (subject and object) (*cf.* Figure 1). The predicate is known as the *property* of a triple, whereas the subject and object may be referred to as *entities* or *concepts*.

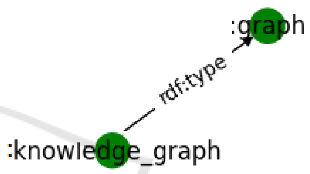


Figure 1: RDF triple represented as a directed graph in a visual representation.

KGs may be described in textual files often encoded in RDF-based languages, such as *Terse RDF Triple Language*[2], also known as *Turtle*, or *TTL*.

In formal terms, a Knowledge Graph $\mathcal{KG} = (\mathcal{V}, \mathcal{E})$ can be represented as a directed graph (digraph) containing a set of vertices $\mathcal{V}$ and directed edges $\mathcal{E}$. Vertices represent entities or concepts, and edges express how such concepts and entities relate to each other. An RDF triple refers to a data entity composed of a subject ($s$), predicate ($p$), and an object ($o$), represented as $t = (s, p, o)$. In KGs, the edges are, then, a set of predicates, such that $\mathcal{E} = \{p_0, p_1, \ldots, p_n\}$. Vertices are, in turn, a set of subjects and objects, such that $\mathcal{V} = \{s_0, s_1, \ldots, s_n, o_0, o_1, \ldots, o_n\}$. A KG may, therefore, be represented as a set of RDF triples, such that, $\mathcal{KG} = \{t_0, t_1, \ldots, t_n\}$, where $t_0 = (s_0, p_0, o_0), t_1 = (s_1, p_1, o_1), \ldots, t_n = (s_n, p_n, o_n)$. A predicate $p_i$ in a triple $t_i = (s_i, p_I, o_i)$ is represented as a directed edge from the subject $s_i$ to the object $o_i$.

### 2.1.1 Temporal Knowledge Graphs

As knowledge evolves, KGs, likewise, may evolve. Considering an initial version of a KG, we have a fi-

[2]https://www.w3.org/TeamSubmission/turtle/ (As of Aug. 2023).

nite set of triples. As such KG changes, we may have at a future time, a different set of triples than we had in its initial version. As possible changes, new triples might have been added to the original set. Also, a subset of triples might have been removed from the original set. We consider a *Temporal Knowledge Graph* (TKG) as a graph that represents not only knowledge in terms of entities and relationships but also, encodes how they change over time.

A Temporal Knowledge Graph $\mathcal{TKG} = \{\mathcal{KG}^0, \mathcal{KG}^1, \cdots, \mathcal{KG}^m\}$, where $\mathcal{KG}^i$ is a KG in which the triples represent facts that are available in a determined time frame $i$, *i.e.*, $\mathcal{KG}^i = \{t_0^i, t_1^i, \cdots, t_n^i\}$. In this sense, if we consider KGs in two different time frames ($i$ and $i+1$), *i.e.*, $\mathcal{KG}^i = \{t_0^i, t_1^i, \cdots, t_n^i\}$, and $\mathcal{KG}^{i+1} = \{t_0^{i+1}, t_1^{i+1}, \cdots, t_n^{i+1}\}$, we may have triples like $t_a \in \mathcal{KG}^i$ and $t_a \in \mathcal{KG}^{i+1}$, meaning that the fact described by $t_a$ is available in both time frames. We may have triples like $t_b \in \mathcal{KG}^i$ and $t_b \notin \mathcal{KG}^{i+1}$, *i.e.*, the fact described by the $t_b$ is available only in the time frame $i$, and not in $i+1$.

### 2.2 Complex Networks

Complex networks can be modeled as large graphs (Latapy and Magnien, 2008). More specifically (Rodrigues, 2019), a complex network can be represented as a Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of vertices $\mathcal{V} = \{v_0, v_1, ..., v_n\}$, representing the nodes of the network, and $\mathcal{E} = \{e_0, e_1, ..., e_m\}$ a set of edges, representing the connections between the nodes of the network, *i.e.*, $e_k = (v_i, v_j)$.

A network can be undirected, *i.e.*, every pair of edges connecting $v_i$ to $v_j$ also connects $v_j$ to $v_i$, *i.e.*, $(v_i, v_j) = (v_j, v_i)$. Also, a network can be directed, represented as a digraph (similarly to KGs), where $(v_i, v_j) \neq (v_j, v_i)$. Network edges can also have weights that indicate the interaction strength between two nodes. In such cases, they can be represented by weighted graphs. Since KGs do not weigh their edges, weighted networks are out of this study's scope.

Real-world networks often present characteristics (da F. Costa et al., 2007) that are not observed when considering random connectivity between the nodes (Erdős et al., 1960). Such characteristics are, for instance, (1) community structures (Girvan and Newman, 2002), *i.e.*, the network nodes can be grouped into sets, which are internally densely connected; (2) the "small-world" phenomenon, which refers to the fact that the average number of edges between nodes is small, while the clustering coefficient is large, as described by the small-world network model (Watts and Strogatz, 1998); and (3) the availability of hubs and power-law degree distributions, as described by the scale-free network model (Albert and Barabási, 2002).

## 3 RELATED WORK

We searched the literature for studies on complex networks and KGs, considering complex network characteristics. In this section, we present a summary of the studies found adhering to such conditions.

The work by (Lü et al., 2022) reviews studies involving complex networks and KGs. Their work introduced a framework for modeling knowledge based on complex networks for usage in conjunction with deep learning models.

The influence of topology on KGs is investigated in the study by (Dörpinghaus et al., 2022). They evaluated the impact of adding extra layers of nodes in KGs generated following the scale-free and small-world models. They compared the relevance of nodes obtained through degree (Freeman, 1978) and betweenness (Brandes, 2001) centralities, before and after the addition of extra nodes. Also focusing on topological aspects, the investigation by (Magnanimi et al., 2023) presented a study concerning the effectiveness of updating portions of KGs. They compared the results obtained considering distinct properties of KGs, including their topology. They considered large KGs following the scale-free, small-world, and even random KGs.

(Chen et al., 2022) combined KGs and other inputs to construct what they call "tripartite" graphs, used for their recommendation method. They observed such tripartite graphs present properties of scale-free networks. They benefit from properties, such as hubs, when embedding their graphs, hence, improving their recommendation method.

The work by Mantle (Mantle et al., 2019), in turn, explored the reasoning over large-scale databases, in special KGs. Although not the focus of their study, they observed in their evaluation the tendency of KGs presenting a scale-free topology, typified by a small number of hubs connected to many nodes, and a large number of nodes with few connections.

Different from the studies found in the literature, our present investigation emphasizes specifically observing complex network properties in widely known real-world KGs. We seek for characteristics that are observed in models that describe real-world complex networks. Our ultimate objective with our experimental procedure is to assess the use of complex network measurements (most notably, centrality metrics) on KGs. To the best of our knowledge, no study in the literature has performed such an analysis.

# 4 METHOD

We considered two real-world complex network models and their characteristics. Considering their definition, we aim to observe whether such characteristics can be found in KGs. Figure 2 illustrates the observed characteristics and their best-fitting models. In Section 4.1 and Section 4.2, we show how a KG can satisfy the properties of such two specific complex network models.

## 4.1 Scale-Free Model

Scale-free networks present a degree distribution asymptotically following a power-law. Such a law denotes that most nodes in the network have a low amount of links, while a few important nodes hold a higher amount of network links. Those are, therefore, the main characteristics of the scale-free model, *i.e.*, degree distribution following power-law, and the presence of hubs.

At this stage, we describe how we observe such scale-free model characteristics on KGs, as scale-free digraphs (Bollobas et al., 2003).

### 4.1.1 Degree Distribution

Degree distributions relate the degrees ($k$) of the nodes in a network with the frequencies ($p_k$) in which they are observed, *i.e.*, $p_k \sim k^{-y}$ (where $y$ is the degree exponent). The degree distribution of scale-free networks follows a power-law distribution, in which we observe a high frequency of low-degree nodes and a low frequency of high-degree nodes. This can be observed when plotting the distribution in a log-log scale, in which the data points can be roughly approximated to a straight line, *i.e.*, $\log p_k \sim -y \log k$, where $\log p_k$ is expected to be linearly dependent on $\log k$ (Barabási and Pósfai, 2016).

KGs are digraphs; each node has both an in-degree ($k_{in}$) and an out-degree ($k_{out}$), *i.e.*, the number of edges pointing towards and away from them, respectively. This way, the degree of a KG node is $k = k_{in} + k_{out}$. When adding a directed edge from node $i$ to node $j$ in a KG, we expect to increase $k^i_{out}$ and $k^j_{in}$. In this sense, we can distinguish two distributions for KGs: in-degree and out-degree distributions, where, similarly, we expect $\log p_{k_{in}} \sim -y \log k_{in}$ and $\log p_{k_{out}} \sim -y \log k_{out}$

### 4.1.2 Hubs

In networks with power-law degree distribution, most nodes present only a few links, while a few other nodes concentrate the majority of links in the network. Such few nodes, called hubs, hold the network together by linking most less-linked nodes.

Given the digraph nature of KGs, where incoming and outgoing edges are available, we can distinguish highly-linked nodes as hubs and authorities (Kleinberg, 1999). In this domain, hubs are nodes with more outgoing links, while authorities are those with more incoming links. We may observe hubs at the tail of the in-degree distribution, as such region denotes the few available nodes with the higher in-degree values. On the other hand, authorities can be analogously observed at the tail portion of out-degree distributions.

## 4.2 Small-World Model

In small-world networks, despite not being direct neighbors, most nodes can reach the majority of other nodes through a small number of steps (*i.e.*, a small path exists between such nodes). On such networks, the degree distribution follows Poisson's distribution. They display a short average path length and a high average clustering coefficient. We describe how we observe such small-world model characteristics on KGs.

### 4.2.1 Degree Distribution and Hubs

On small-world networks, the degree distribution follows Poisson's distribution. For this reason, we observe that most of the nodes have an average degree. This also denotes an overabundance of hubs.

For KGs, we distinguish the degree distribution in two distinct distributions (*cf.* Section 4.1.1). Therefore, we may observe the patterns on $p_{k_{in}}$ vs $k_{in}$ and $p_{k_{out}}$ vs $k_{out}$ plots (including the overabundance of hubs).

### 4.2.2 Average Short Path Length

Small-world networks present a small average short path length, denoting the connectivity between most nodes and that a small path exists between two distinct nodes in the network. Considering a network of $N$ nodes, the average short path length ($< d >$) is given by $< d >= \frac{1}{N(N-1)} \sum_{i \neq j}^{N} d(i, j)$, where $d(i, j)$ is the shortest path length between nodes $i$ and $j$. If the network contains disconnected components, then $< d >$ cannot be calculated due to distances between some nodes diverging to infinity.

In KGs, the path must consider the direction of the edges. In this sense, we could have paths in which we may reach node $j$ from node $i$. However, we may not reach node $j$ from node $i$, *i.e.*, $d(i, j) \neq d(j.i)$.

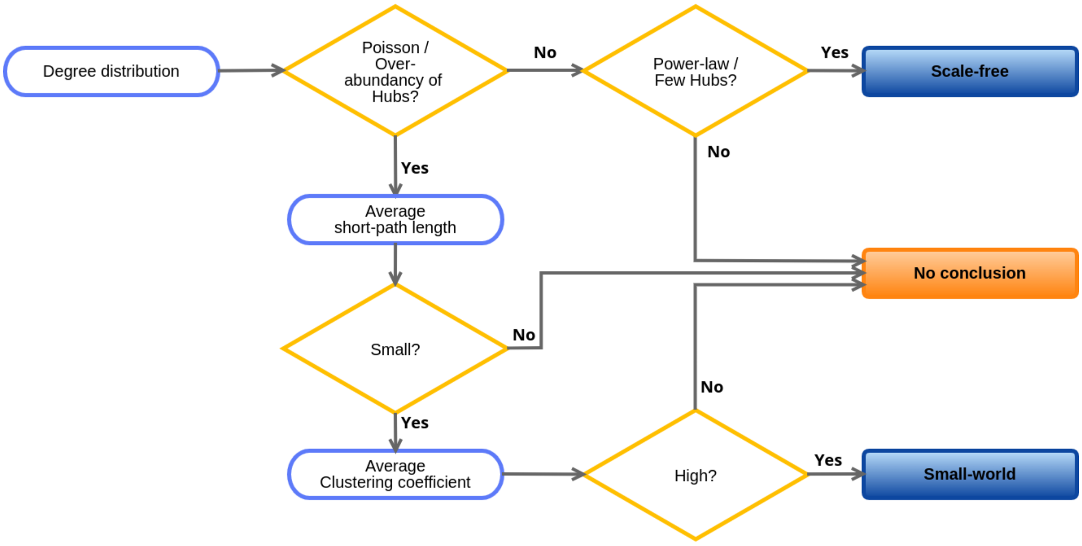


Figure 2: **Complex Network properties' assessment**. We observe such characteristics on KGs to verify the best-fitting model.

Furthermore, the KG must be strongly connected to calculate the average shortest path length.

### 4.2.3 Average Clustering Coefficient

Small-world networks present a high average clustering coefficient. The clustering coefficient ($C_i$) of a node $i$ denotes how linked to each other its neighbors are. The value ranges from 0 to 1, representing no connectivity and full connectivity, respectively. The average value ($<C>$) computes the metric for all the nodes in the network. It is defined as $<C>=\frac{1}{N}\sum_{i=1}^{N}C_i$.

For KGs and their possibility of incoming and outgoing edges between the nodes, to consider complete connectivity, there should be at least one incoming and an outgoing edge between a node's neighbors.

## 5 EVALUATION

We performed an analysis to assess if widely used KGs present characteristics that are observed in real-world complex networks rather than random networks. We considered the characteristics of two models of complex networks: Scale-free and Small-world (*cf.* Section 4). We computed from the KGs, their degree distributions, and the availability of hubs. Also, we sought to compute their average short path lengths and average clustering coefficients.

We conducted such evaluation[3] on subgraphs of two well-known widely-used KGs: *DBpedia* (*cf.* Section 5.1) and *Wikidata* (*cf.* Section 5.2). We took two portions of such knowledge bases (*i.e.*, sub-KGs), from which centrality measurements were extracted from studies found in the literature, aiming for distinct objectives. In one study (Kalloubi et al., 2016), a sub-KG comprising all the relationships about the *Oracle corporation* entity, from *DBpedia* was used. In another one (Puspa Rinjeni et al., 2022), relationships from the *Movie* entity, from both *DBpedia*[4][5] and *Wikidata*[6][7] were used. In our procedure, we first observed the characteristics of each of the sub-KGs alone, and then we merged both into a single KG, to confirm if the same characteristics were still observed. The first and second sub-KGs represent the relationships of the same chosen entities on both knowledge bases. Figure 3 illustrates the adopted procedure.

Finally, we considered observing the same characteristics on a TKG (*cf.* Section 5.3), to verify if such real-world complex networks' characteristics can also change over time. The TKG used in this analysis was generated in a recent study (Rossanez et al., 2020) for centrality measurement extraction. In the following, we present the observed characteristics in all the aforementioned KGs.

[3]https://github.com/rossanez/complexnw-kg (As of Aug. 2023).

[4]https://dbpedia.org/data/Oracle_Corporation.ttl (As of Aug. 2023).

[5]https://dbpedia.org/data/Movie.ttl (As of Aug. 2023).

[6]https://www.wikidata.org/wiki/Special:EntityData/Q19900.ttl (As of Aug. 2023).

[7]https://www.wikidata.org/wiki/Special:EntityData/Q11424.ttl (As of Aug. 2023).

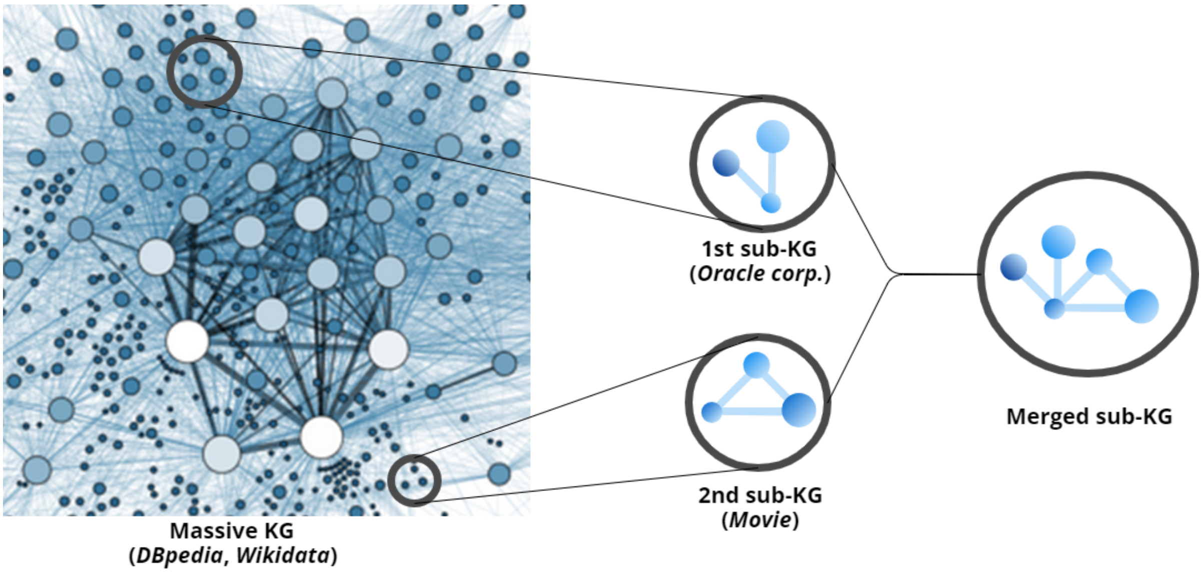


Figure 3: **KG creation procedure**. Applied to *DBpedia* and *Wikidata*. We took two portions of each database (*i.e.*, the 1st and 2nd sub-KGs), corresponding to all the relationships about the *Oracle corporation*, and *Movie* entities, respectively.

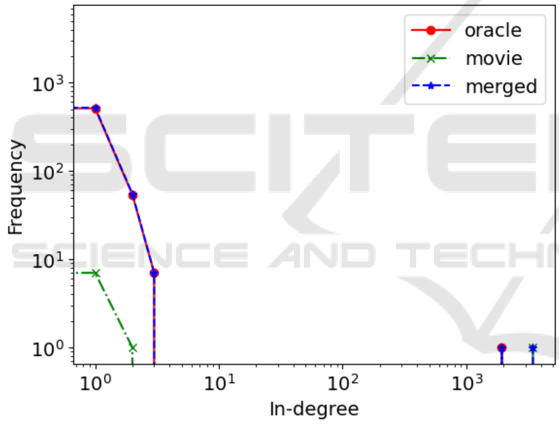


Figure 4: **DBpedia in-degree distributions**. All sub-KGs present a power-law pattern.

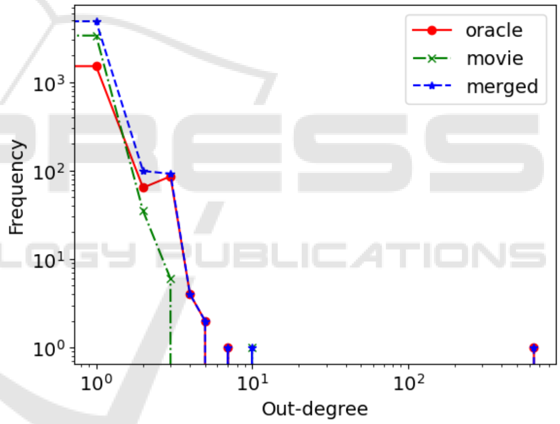


Figure 5: **DBpedia out-degree distributions**. All sub-KGs present a power-law pattern.



### 5.1 Results on the DBpedia Analysis

Figure 4 shows the in-degree distributions for the three sub-KGs retrieved from *DBpedia*.

We observe that all cases follow a similar distribution, where most nodes present a low in-degree, while few concentrate higher degrees. The behavior resembles a power-law distribution. Figure 5 shows the out-degree distributions for the same sub-KGs, and we observe similar patterns.

Another aspect evidenced by a few higher-in-degree nodes is that, although not over-abundant, we found the presence of hubs, as observed in the tail portion of the distribution illustrated by Figure 4.

Table 1 presents computed metrics obtained for the KGs under analysis. We observe in all cases that the density of the KGs is very small. Considering the number of available nodes, the number of observed edges is too small when compared to the number of directed edges that would be necessary for a complete digraph (*i.e.*, two edges – one in each direction – for each pair of available edges).

In addition, such graphs are not strongly connected. For this reason, it is not possible to calculate the average short path length. As those directed graphs are sparse, reaching all nodes starting from a randomly given node is impossible. A small average short path length would be necessary in the small-world network model. Another condition for such a model would be a high average clustering coefficient.

Table 1: **Metrics for DBpedia KGs**. Number of nodes (N), edges (E), density (D), average short path length ($< d >$), and average clustering coefficient ($< C >$).

| | N | E | D | $< d >$ | $< C >$ |
|---|---|---|---|---|---|
| Oracle | 2143 | 2577 | 0.00056 | N/A | 0.0 |
| Movie | 3411 | 3460 | 0.00029 | N/A | 0.0 |
| Merged | 5554 | 6037 | 0.00019 | N/A | 0.0 |

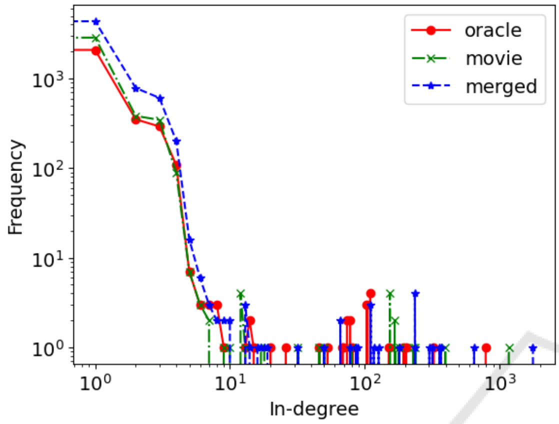


Figure 6: **Wikidata in-degree distribution**. All sub-KGs present a power-law pattern.

Table 1 presents the KGs have practically zero clustering coefficient, also explained by the very low number of edges for the available nodes.

Such KGs, therefore, better fit the scale-free network model due to the observed power-law degree distribution and presence of hubs, rather than the small-world model, as they do not have a small average short path length and small clustering coefficient. The KGs are not random and display characteristics of real-life complex networks.

## 5.2 Results on the Wikidata Analysis

Figure 6 presents the degree distributions for the three sub-KGs retrieved from *Wikidata*.

Similarly to what we observed from the results regarding *DBpedia* analysis, all three KGs' degree distributions resemble a power-law distribution. We indicate the presence of hubs at the tail of the in-degree distribution. Figure 7 shows the out-degree distributions for the same sub-KGs, and we observe similar patterns.

Table 2 presents the metrics obtained for the *Wikidata* KGs. The density of the KGs, although still small, is slightly higher when compared to *DBpedia*. We found a higher amount of edges for the given nodes available (*i.e.*, *Wikidata* KGs have more triples than those of *DBpedia*).

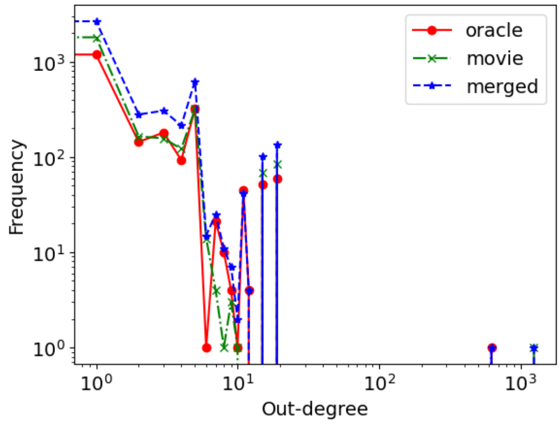


Figure 7: **Wikidata out-degree distribution**. All sub-KGs present a power-law pattern.

Table 2: **Metrics for Wikidata KGs**. Number of nodes (N), edges (E), density (D), average short path length ($< d >$), and average clustering coefficient ($< C >$).

| | N | E | D | $< d >$ | $< C >$ |
|---|---|---|---|---|---|
| Oracle | 3052 | 7275 | 0.00078 | N/A | 0.00806 |
| Movie | 4069 | 8741 | 0.00052 | N/A | 0.00721 |
| Merged | 6530 | 15051 | 0.00035 | N/A | 0.00692 |

The KGs are similarly not strongly connected, so their average shortest path lengths cannot be determined. They present a higher clustering coefficient than *DBpedia's*, although still very small (the value ranges from 0 to 1). The higher amount of triples/edges available can also explain the increase.

Similar to *DBpedia*, such KGs are a better fit for the scale-free network model rather than the small-world model. We can, therefore, also indicate these KGs are not random and display properties of real-world complex networks.

## 5.3 Results on the Temporal KG Analysis

We considered three temporal instances of a TKG generating from abstracts of an annual scientific event, more specifically, representing the editions of 2019, 2020, and 2021 of the International Semantic Web Conference[8] (ISWC). Figure 8 presents the in-degree distributions of the three referred temporal instances.

Similar to the other observed cases, they follow a power-law distribution pattern, evidencing the presence of hubs. Figure 9 shows the out-degree distributions for the same sub-KGs. Table 3 shows that the

[8] https://link.springer.com/conference/semweb (As of Aug. 2023).

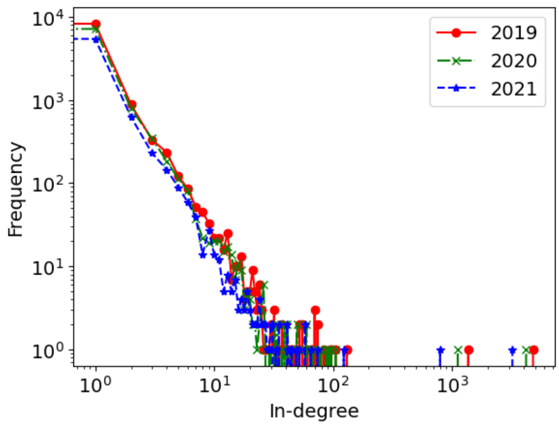


Figure 8: **TKG in-degree distribution**. All three instances present a power-law pattern.

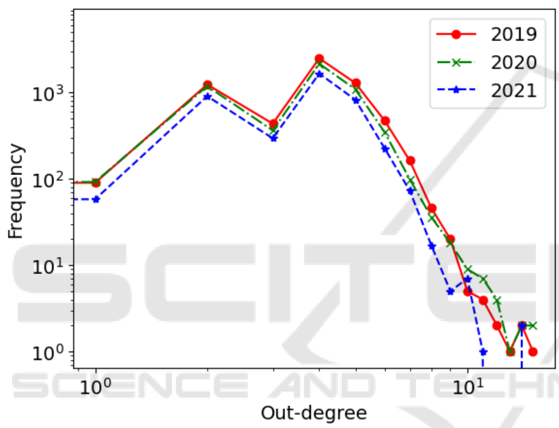


Figure 9: **TKG out-degree distribution**. All three instances present a power-law pattern.

Table 3: **Metrics for TKG**. Number of nodes (N), edges (E), density (D), average short path length ($< d >$), and average clustering coefficient ($< C >$).

| | N | E | D | $< d >$ | $< C >$ |
|---|---|---|---|---|---|
| 2019 | 12285 | 24984 | 0.00016 | N/A | 0.02860 |
| 2020 | 10618 | 21147 | 0.00018 | N/A | 0.02932 |
| 2021 | 8003 | 15672 | 0.00024 | N/A | 0.02923 |

temporal instances present a small density. However, it shows larger edges (triples) for the available nodes.

Like the other cases, neither of the temporal instances is strongly connected, so the average short path length cannot be calculated. They also present a small average clustering coefficient. Following the same outcome of the previously analyzed KGs, the temporal instances of the TKG fit the scale-free network model and not the small-world model. In addition, we found that the characteristics are maintained over time, allowing us to state that the TKG is not random and displays characteristics of real-world complex networks. It is worth mentioning that no substantial changes in complex network patterns were observed in the temporal evolution.

## 6 DISCUSSION

The results presented in Section 5 indicated that real-world KGs present characteristics of complex network models in KGs. While, in theory, complex networks may be represented by random graphs (Erdős et al., 1960), models describe real-world complex networks (Watts and Strogatz, 1998; Albert and Barabási, 2002) presenting characteristics that are not observed when considering random behavior. Most notably, some of such characteristics are power-law degree distribution, presence of hubs, high clustering coefficient, etc.

To observe such characteristics in KGs, we had to refer to their definition, considering a digraph that allows multiple parallel edges. To address our first research question (*i.e.*, **RQ1**), we then had to transpose the expected properties from non-directed graphs that represent complex networks to the KG domain. Our results on real-world KGs indicate their tendency to follow the scale-free model, in which the degree distributions, more specifically of both in-degree and out-degrees, follow a power-law. This means that on real-world KGs, we are expected to encounter few nodes with higher degree nodes, and many nodes with low degree. This, in turn, indicates the presence of not-over-abundant hubs, represented by the fewer nodes with the highest degrees.

While presenting a good fit for scale-free models, real-life KGs did not present the characteristics expected for the small-world model. We observed that real-life KGs have low density, *i.e.*, and have a small number of edges, in comparison with their number of nodes. Furthermore, real-life KGs often present disjoint portions, making them not strongly connected. Considering such aspects, we expect a low average clustering coefficient and a high average short path length. Despite not fitting such a model, we did observe a tendency in all evaluated KGs to follow the scale-free model. For this reason, we can positively answer **RQ2**.

We chose to include a TKG in our study, as centrality measurements are employed to characterize the knowledge evolution over TKGs in the literature (Rossanez et al., 2020). The analyzed TKG presented the same characteristics in all the temporal instances. The power-law degree distribution and presence of hubs were consistently observed, as well as a small clustering coefficient and not-strongly connectedness. All instances, therefore, have shown a good fit for the scale-free model.

Of course, the KGs observed in this study do not correspond to the totality of the datasets they were obtained from. Despite our efforts to observe distinct portions of them, which hindered the same results, there is a possibility that a random portion not covered by this study might fail to present the same results. We could have considered more large-scale KGs besides *DBpedia* and *Wikidata*, to further assure our findings. Other complex network characteristics, such as, for instance, the availability of communities (Girvan and Newman, 2002), can be explored, as well as other models that describe such networks (Anderson and Dragićević, 2020). On the other hand, the obtained results suggest that, indeed, KGs hold relevant properties of complex networks.

With such a statement, we, thus, positively answer our final research question (*i.e.*, **RQ3**). Complex networks measurements, especially centrality metrics, can, therefore, be used with confidence in KG-based analysis, as already being done in several research studies (Dörpinghaus et al., 2022; Park et al., 2019; Rossanez et al., 2020; Sadeghi et al., 2021; Tilly and Livan, 2021).

## 7 CONCLUSION

This study investigated the complex network properties found in well-known and recently used knowledge graphs. To the best of our knowledge, this is the first study focusing on demonstrating those properties. Performed evaluations involving the *DBpedia* and the *Wikidata* knowledge graphs to confirm their complex network properties, therefore validating existing studies dedicated to the use of complex network measurements in the characterization of knowledge graphs (e.g., relevance of concepts) (Rossanez et al., 2020). Future work encompasses the analysis of the temporal evolution of complex network properties found in relevant knowledge graphs and the connection of those properties with the effectiveness of typical reasoning algorithms. We plan to investigate the creation of synthetic knowledge graphs and their use for training machine learning algorithms. The creation of those synthetic datasets would comply with pre-defined network properties (Dadauto et al., 2023).

## ACKNOWLEDGEMENTS

This work was supported by the São Paulo Research Foundation (FAPESP) (Grant #2022/15816-5)[9].

## REFERENCES

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97.

Anderson, T. and Dragićević, S. (2020). Complex spatial networks: Theory and geospatial applications. *Geography Compass*, 14(9):e12502.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., and Cudré-Mauroux, P., editors, *The Semantic Web*, pages 722–735, Berlin, Heidelberg. Springer Berlin Heidelberg.

Barabási, A.-L. and Pósfai, M. (2016). *Network science*. Cambridge University Press, Cambridge.

Bollobas, B., Borgs, C., Chayes, J., and Riordan, O. (2003). Directed scale-free graphs. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 132–139.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2):163–177.

Candan, K. S., Liu, H., and Suvarna, R. (2001). Resource description framework: Metadata and its applications. *SIGKDD Explor. Newsl.*, 3(1):6–19.

Chen, Y., Yang, M., Zhang, Y., Zhao, M., Meng, Z., Hao, J., and King, I. (2022). Modeling scale-free graphs with hyperbolic geometry for knowledge-aware recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, WSDM '22, page 94–102, New York, NY, USA. Association for Computing Machinery.

da F. Costa, L., Rodrigues, F. A., Travieso, G., and Boas, P. R. V. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242.

Dadauto, C. V., da Fonseca, N. L. S., and da Silva Torres, R. (2023). Data-driven intra-autonomous systems graph generator. *CoRR*, abs/2308.05254.

Dörpinghaus, J., Weil, V., Düing, C., and Sommer, M. W. (2022). Centrality measures in multi-layer knowledge graphs. In *Communication Papers of the 17th Conference on Computer Science and Intelligence Systems*, page 163–170.

Ehrlinger, L. and Wöß, W. (2016). Towards a definition of knowledge graphs. In *12th International Conference*

[9]The opinions expressed in this work do not necessarily reflect those of the funding agencies.

*on Semantic Systems (SEMANTiCS2016)*, pages 14–17.

Erdős, P., Rényi, A., et al. (1960). On the evolution of random graphs. *Publ. math. inst. hung. acad. sci*, 5(1):17–60.

Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., and Vrandečić, D. (2014). Introducing wikidata to the linked data web. In Mika, P., Tudorache, T., Bernstein, A., Welty, C., Knoblock, C., Vrandečić, D., Groth, P., Noy, N., Janowicz, K., and Goble, C., editors, *The Semantic Web – ISWC 2014*, pages 50–65, Cham. Springer International Publishing.

Faerber, M., Bartscherer, F., Menne, C., and Rettinger, A. (2017). Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic Web*, 9:1–53.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1:215–239.

Girvan, M. and Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826.

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., and Zimmermann, A. (2021). Knowledge graphs. *ACM Comput. Surv.*, 54(4).

Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S. (2022). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514.

Kalloubi, F., Nfaoui, E. H., and Beqqali, O. E. (2016). On using graph centrality measures for dbpedia-based tweet entity linking. In *2016 International Conference on Information Technology for Organizations Development (IT4OD)*, pages 1–7.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632.

Latapy, M. and Magnien, C. (2008). Complex network measurements: Estimating the relevance of observed properties. In *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*, pages 1660–1668.

Lü, J., Wen, G., Lu, R., Wang, Y., and Zhang, S. (2022). Networked knowledge and complex networks: An engineering view. *IEEE/CAA Journal of Automatica Sinica*, 9(8):1366–1383.

Magnanimi, D., Bellomarini, L., Ceri, S., and Martinenghi, D. (2023). Reactive company control in company knowledge graphs. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 3336–3348.

Mantle, M., Batsakis, S., and Antoniou, G. (2019). Large scale distributed spatio-temporal reasoning using real-world knowledge graphs. *Knowledge-Based Systems*, 163:214–226.

Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., and Taylor, J. (2019). Industry-scale knowledge graphs: Lessons and challenges: Five diverse technology companies show how it's done. *Queue*, 17(2):48–75.

Park, N., Kan, A., Dong, X. L., Zhao, T., and Faloutsos, C. (2019). Estimating node importance in knowledge graphs using graph neural networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 596–606, New York, NY, USA. Association for Computing Machinery.

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

Puspa Rinjeni, T., Suci Indasari, S., Indriawan, A., and Aini Rakhmawati, N. (2022). Movies analysis on dbpedia and wikidata using community detection and centrality algorithms. In *2022 International Electronics Symposium (IES)*, pages 380–386.

Rodrigues, F. A. (2019). *Network Centrality: An Introduction*, pages 177–196. Springer International Publishing, Cham.

Rossanez, A., dos Reis, J. C., and da Silva Torres, R. (2020). Representing scientific literature evolution via temporal knowledge graphs. In *6th Managing the Evolution and Preservation of the Data Web (MEPDaW) Workshop, International Semantic Web Conference (ISWC)*, pages 33–42.

Sadeghi, A., Collarana, D., Graux, D., and Lehmann, J. (2021). Embedding knowledge graphs attentive to positional and centrality qualities. In Oliver, N., Pérez-Cruz, F., Kramer, S., Read, J., and Lozano, J. A., editors, *Machine Learning and Knowledge Discovery in Databases. Research Track*, pages 548–564, Cham. Springer International Publishing.

Tilly, S. and Livan, G. (2021). Macroeconomic forecasting with statistically validated knowledge graphs. *Expert Systems with Applications*, 186:115765.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

Zou, X. (2020). A survey on application of knowledge graph. *Journal of Physics: Conference Series*, 1487(1):012016.