Characterizing the Hypergraph-of-Entity Representation Model

José Devezas and Sérgio Nunes

INESC TEC and Faculty of Engineering, University of Porto
Rua Dr. Roberto Frias, s/n, 4200-465, Porto, Portugal
{jld, ssn}@fe.up.pt

Abstract. The hypergraph-of-entity is a joint representation model for terms, entities and their relations, used as an indexing approach in entity-oriented search. In this work, we characterize the structure of the hypergraph, from a microscopic and macroscopic scale, as well as over time with an increasing number of documents. We use a random walk based approach to estimate shortest distances and node sampling to estimate clustering coefficients. We also propose the calculation of a general mixed hypergraph density based on the corresponding bipartite mixed graph. We analyze these statistics for the hypergraph-of-entity, finding that hyperedge-based node degrees are distributed as a power law, while node-based node degrees and hyperedge cardinalities are log-normally distributed. We also find that most statistics tend to converge after an initial period of accentuated growth in the number of documents.

Keywords: hypergraph-of-entity combined data, indexing, representation model, hypergraph analysis, characterization

1 Introduction

Complex networks have frequently been studied as graphs, but only recently has attention been given to the study of complex networks as hypergraphs [11]. The hypergraph-of-entity [10] is a hypergraph-based model used to represent combined data [4, §2.1.3]. That is, it is a joint representation of corpora and knowledge bases, integrating terms, entities and their relations. It attempts to solve, by design, the issues of representing combined data through inverted indexes and quad indexes. The hypergraph-of-entity, together with its random walk score [10, §4.2.2], is also an attempt to generalize several tasks of entity-oriented search. This includes ad hoc document retrieval and ad hoc entity retrieval, as well as the recommendation-alike tasks of related entity finding and entity list completion. However, there is a tradeoff. One one side, the random walk score acts as a general ranking function. On the other side, it performs below traditional baselines like TF-IDF. Since ranking is particularly dependent on the structure of the hypergraph, a characterization is a fundamental step towards improving the representation model and, with it, the retrieval performance.

Accordingly, our focus was on studying the structural properties of the hypergraph. This is a task that presents some challenges, both from a practical sense and from a theoretical perspective. While there are many tools [9,5] and formats [17,8] for the analysis and transfer of graphs, hypergraphs still lack clear frameworks to perform these functions, making their analysis less trivial. Even formats like GraphML [8] only support undirected hypergraphs. Furthermore, there is still an ongoing study of several aspects of hypergraphs, some of which are trivial in graph theory. For example, the adjacency matrix is a well-established representation of a graph, however recent work is still focusing on defining an adjacency tensor for representing general hypergraphs [21]. As a scientific community, we have been analyzing graphs since 1735 and, even now, innovative ideas in graph theory are still being researched [1]. However, hypergraphs are much younger, dating from 1970 [6], and thus there are still many open challenges and contribution opportunities.

In this work, we take a practical application of hypergraphs, the hypergraph-of-entity, as an opportunity to establish a basic framework for the analysis of hypergraphs. In Section 2, we begin by providing an overview on the analysis of the inverted index, knowledge bases and hypergraphs, covering the three main aspects of the hypergraph-of-entity. In Section 3, we describe our characterization approach, covering shortest distance estimation with random walks and clustering coefficient estimation with node sampling, as well as proposing a general mixed hypergraph density formula by establishing a parallel with the corresponding bipartite mixed graph. In Section 4, we present the results of a characterization experiment of the hypergraph-of-entity for a subset of the INEX 2009 Wikipedia collection and, in Section 5, we present the conclusions and future work.

2 Reference Work

The hypergraph-of-entity is a representation model for indexing combined data, jointly modeling unstructured textual data from corpora and structured interconnected data from knowledge bases. As such, before analyzing a hypergraph from this model, we surveyed existing literature on inverted index analysis, as well knowledge base analysis. We then surveyed literature specifically on the analysis of hypergraphs, particularly focusing on statistics like the clustering coefficient, the shortest path lengths and the density.

Analyzing Inverted Indexes There are several models based on the inverted index that combine documents and entities [7,3] and that are comparable with the hypergraph-of-entity. There has also been work that analyzed the inverted index, particularly regarding query evaluation speed and space requirements [23,25].

Voorhees [23] compared the efficiency of the inverted index with the top-down cluster search. She analyzed the storage requirements of four test collections, measuring the total number of documents and terms, as well as the average number of terms per document. She then analyzed the disk usage per collection, measuring the number of bytes for document vectors and the inverted index. Finally, she measured CPU time in number of instructions and the I/O time in number of data pages accessed at least once, also including the query time in seconds.

Zobel et al. [25] took a similar approach to compare the inverted index and signature files. First, they characterized two test collections, measuring size in megabytes, number of records and distinct words, as well as the record length, and the number of words, distinct words and distinct words without common terms per record. They also analyzed disk space, memory requirements, ease of index construction, ease of update, scalability and extensibility.

For the hypergraph-of-entity characterization, we do not focus on measuring efficiency, but rather on studying the structure and size of the hypergraph.

Analyzing Knowledge Bases Studies have been made to characterize the entities and triples in knowledge bases. In particular, given RDF's graph structure, we are interested in understanding which statistics are relevant for instance to discriminate between the typed nodes.

Halpin [16] took advantage of Microsoft's *Live.com* query log to reissue entity and concept queries over their FALCON-S semantic web search engine. They then studied the results, characterizing their source, triple structure, RDF and OWL classes and properties, and the power-law distributions of the number of URIs, both returned as results and as part of the triples linking to the results. They focused mostly on measuring the frequency of different elements or aggregations (e.g., top-10 domain names for the URIs, most common data types, top vocabulary URIs).

Ge et al. [14] defined an object link graph based on the graph induced by the RDF graph, based on paths linking objects (or entities), as long as they are either direct or established through blank nodes. They then studied this graph for the Falcons Crawl 2008 and 2009 datasets (FC08 and FC09), which included URLs from domains like bio2rdf.org or dbpedia.org. They characterized the object link graph based on density, using the average degree as an indicator, as well as connectivity, analyzing the largest connected component and the diameter. They repeated the study for characterizing the structural evolution of the object link graph, as well its domain-specific structures (according to URL domains). Comparing two snapshots of the same data enabled them to find evidence of the scale-free nature of the network. While the graph almost doubled in size from FC08 to FC09, the average degree remained the same and the diameter actually decreased.

Fernandez et al. [12] focused on studying the structural features of RDF data, identifying redundancy through common structural patterns, proposing several specific metrics for RDF graphs. In particular, they proposed several subject and object degrees, accounting for the number of links from/to a given subject/object (outdegree and indegree), the number of links from a $\langle$*subject*, *predicate*$\rangle$ (partial outdegree) and to a $\langle$*predicate*, *object*$\rangle$ (partial indegree), the number of distinct predicates from a subject (labeled outdegree) and to an object (labeled indegree), and the number of objects linked from a subject through a single predicate (direct outdegree), as well as the number of subjects linking to an object through a single predicate (direct indegree). They also measured predicate degree, outdegree and indegree. They proposed common ratios to account for shared structural roles of subjects, predicates and objects (e.g., subject-object ratio). Global metrics were also defined for measuring the maximum and average outdegree of subject and object nodes for the whole graph. Another analysis approach was focused on the predicate lists per subject, measuring the ratio of repeated lists and their degree, as well as the number of lists per predicate. Finally, they also defined several statistics to measure typed subjects and classes, based on the *rdf:type* predicate.

While we study a hypergraph that jointly represents terms, entities and their relations, we focus on a similar characterization approach, that is more based on structure and less based on measuring performance.

Analyzing Hypergraphs Hypergraphs [6] have been around since the 1970s and, because they are able to capture higher-order relations, there are either conceptually different or multiple counterparts to the equivalent graph statistics. Take for instance the node degree. While graphs only have a node degree, indegree and outdegree, hypergraphs can also have a hyperedge degree, which is the number of nodes in a hyperedge [18]. The hyperedge degree also exists for directed hyperedges, in the form of a tail degree and a head degree[1]. The tail degree is based on the cardinality of the source node set and the head degree is based on the cardinality of the target node set. In this work we will rely on the degree, clustering coefficient, average path length, diameter and density to characterize the hypergraph-of-entity.

Ribeiro et al. [22] proposed the use of multiple random walks to find shortest paths in power law networks. They found that random walks had the ability to observe a large fraction of the network and that two random walks, starting from different nodes, would intersect with a high probability. Glabowski et al. [15] contributed with a shortest path computation solution based on ant colony optimization, clearly structuring it as pseudocode, while providing several configuration options. Parameters included the number of ants, the influence of pheromones and other data in determining the next step, the speed of evaporation of the pheromones, the initial, minimum and maximum pheromone levels, the initial vertex and an optional end vertex. Li [19] studied the computation of shortest paths in electric networks based on random walk models and ant colony optimization, proposing a current reinforced random walk model inspired by the previous two. In this work, we also use a random walk based approach to approximate shortest paths and estimate the average path length and diameter of the graph.

Gallagher and Goldberg [13, Eq.4] provide a comprehensive review on clustering coefficients for hypergraphs. The proposed approach for computing the clustering coefficient in hypergraphs accounted for a pair of nodes, instead of a single node, which is more frequent in graphs. Based on these two-node clustering coefficients, the node cluster coefficient was then calculated. Two-node

[1] Tail and head is used in analogy to an arrow, not a list.

clustering coefficients measured the fraction of common hyperedges between two nodes, through the intersection of the incident hyperedge sets for the two nodes. It then provided different kinds of normalization approaches, either based on the union, the maximum or minimum cardinality, or the square root of the product of the cardinalities of the hyperedge sets. The clustering coefficient for a node was then computed based on the average two-node clustering coefficient for the node and its neighbors.

The codegree Turán density [20] can be computed for a family $\mathcal{F}$ of k-uniform hypergraphs, also known as k-graphs. It is calculated based on the codegree Turán number (the extremal number), which takes as parameters the number of nodes n and the family $\mathcal{F}$ of k-graphs. In turn, the codegree Turán number is calculated based on the minimum number of nodes, taken from all sets of $r-1$ vertices of each hypergraph H that, when united with an additional vertex, will form a hyperedge from H. The codegree density for a family $\mathcal{F}$ of hypergraphs is then computed based on $\limsup_{n\to\infty} \frac{\text{co-ex}(n,\mathcal{F})}{n}$. Since this was the only concept of density we found associated with hypergraphs or, more specifically, a family of k-uniform hypergraphs, we opted to propose our own density formulation (Section 3). The hypergraph-of-entity is a single general mixed hypergraph. In other words, it is not a family of hypergraphs, it contains hyperedges of multiple degrees (it's not k-uniform, but general) and it contains undirected and directed hyperedges (it's mixed). Accordingly, we propose a density calculation based on the counterpart bipartite graph of the hypergraph, where hyperedges are translated to bridge nodes.

3 Characterization Approach

Graphs can be characterized at a microscopic, mesoscopic and macroscopic scale. The microscopic analysis is concerned with statistics at the node-level, such as the degree or clustering coefficient. The mesoscopic analysis is concerned with statistics and patterns at the subgraph-level, such as communities, network motifs or graphlets. The macroscopic analysis is concerned with statistics at the graph-level, such as average clustering coefficient or diameter. In this work, our analysis of the hypergraph is focused on the microscopic and macroscopic scales. We compute several statistics for the whole hypergraph, as well as for snapshot hypergraphs that depict growth over time. Some of these statistics are new to hypergraphs, when compared to traditional graphs. For instance, nodes in directed graphs have an indegree and an outdegree. However, nodes in directed hypergraphs have four degrees, based on incoming and outgoing nodes, as well as on incoming and outgoing hyperedges. While in graphs all edges are binary, leading to only one other node, in hypergraphs hyperedges are n-ary, leading to multiple nodes, and thus different degree statistics. While some authors use 'degree' to refer to node and hyperedge degrees [24, §4][18, §Network Statistics in Hypergraphs], in this work we opted to use the 'degree' designation when referring to nodes and the 'cardinality' designation when referring to hyperedges.

This is to avoid any confusion for instance between an “hyperedge-induced” node degree and a hyperedge cardinality.

For the whole hypergraph, we compute node degree distributions based on nodes and hyperedges, and hyperedge cardinality distributions. For snapshots, we compute average node degrees and hyperedge cardinalities. For both, we compute the estimated clustering coefficient, average path length and diameter, as well as the density and space usage statistics.

Estimating Shortest Distances with Random Walks Ribeiro et al. [22] found that, in power law networks, there is a high probability that two random walk paths, usually starting from different nodes, will intersect and share a small fraction of nodes. We took advantage of this conclusion, adapting it to a hypergraph, in order to compute a sample of shortest paths and their length, used to estimate the average path length and diameter. We considered two (ordered) sets S_1 and S_2 of nodes sampled uniformly at random, each of size $s = |S_1| = |S_2|$. We then launched r random walks of length ℓ from each pair of nodes S_1^i and S_2^i. For a given pair of random walk paths, we iterated over the nodes in the path starting from S_1^i, until we found a node in common with the path starting from S_2^i. At that point, we merged the two paths based on the common node, discarding the suffix of the first path and the prefix of second path. As the number of iterations r increased, we progressively approximated the shortest path for the pair of nodes. This enabled us to generate a sample of approximated shortest path lengths, which could be used to compute the estimated diameter (its maximum) and the estimated average path length (its mean).

Estimating Clustering Coefficients with Node Sampling In a graph, the clustering coefficient is usually computed for a single node and averaged over the whole graph. As shown by Gallagher and Goldberg [13, §I.A.], in hypergraphs the clustering coefficient is computed, at the most atomic level, for a pair of nodes. The clustering coefficient for a node is then computed based on the averaged two-node clustering coefficients between the node and each of its neighbors (cf. Gallagher and Goldberg [13, Eq.4]). Three options were provided for calculating the two-node clustering coefficient, one of them based on the Jaccard index between the neighboring hyperedges of each node [13, Eq.1], which we use in this work.

As opposed to computing it for all nodes, we estimated the clustering coefficients for a smaller sample S of nodes. Furthermore, for each sampled node $s_i \in S$, we also sampled its neighbors $N_S(s_i)$ for computing the two-node clustering coefficients. We then applied the described equations to obtain the clustering coefficients for each node s_i and a global clustering coefficient based on the overall average.

Computing the Density of General Mixed Hypergraphs A general mixed hypergraph is general (or non-uniform) in the sense that its hyperedges can contain an arbitrary number of vertices, and it is mixed in the sense that it can contain hyperedges that are either undirected and directed. We compute a hypergraph's density by analogy with its corresponding bipartite graph, which contains all nodes from the hypergraph, along with connector nodes representing the hyperedges.

Consider the hypergraph $H = (V, E)$, with $n = |V|$ nodes and $m = |E|$ hyperedges. Also consider the set of all undirected hyperedges E_U and directed hyperedges E_D, where $E = E_U \cup E_D$. Their subsets E_U^k and $E_D^{k_1,k_2}$ should also be respectively considered, where E_U^k is the subset of undirected hyperedges with k nodes and $E_D^{k_1,k_2}$ is the subset of directed hyperedges with k_1 tail (source) nodes, k_2 head (target) nodes and $k = k_1 + k_2$ nodes, assuming the hypergraph only contains directed hyperedges between disjoint tail and head sets. This means that the union of $E_U = E_U^1 \cup E_U^2 \cup E_U^3 \cup \dots$ and $E_D = E_D^{1,1} \cup E_D^{1,2} \cup E_D^{2,1} \cup E_D^{2,2} \cup \dots$ forms the set of all hyperedges E. We use it as a way to distinguish between hyperedges with different degrees. This is important because, depending on the degree k, the hyperedge will contribute differently to the density, when considering the corresponding bipartite graph. For instance, one undirected hyperedge with degree $k = 4$ will contribute with four edges to the density. Accordingly, we derive the density of a general mixed hypergraph as shown in Equation 1.

$$D = \frac{2\sum_k k|E_U^k| + \sum_{k_1,k_2}(k_1 + k_2)|E_D^{k_1,k_2}|}{2(n+m)(n+m-1)} \tag{1}$$

In practice, this is nothing more than a comprehensive combination of the density formulas for undirected and directed graphs. On one side, we consider the density of a mixed graph that should result of the combination of an undirected simple graph and a directed simple graph. That is, each pair of nodes can be connected, at most, by an undirected edge and two directed edges of opposing directions. On the other side, we use hypergraph notation to directly obtain the required statistics from the corresponding mixed bipartite graph, thus calculating the analogous density for a hypergraph.

4 Analyzing the Hypergraph-of-Entity

We indexed a subset of the INEX 2009 Wikipedia collection given by the 7,487 documents appearing in the relevance judgments of 10 random topics. We then computed global statistics (macroscale), local statistics (microscale) and temporal statistics. Temporal statistics were based on an increasingly larger number of documents, by creating several snapshots of the index, through a 'limit' parameter, until all documents were considered.

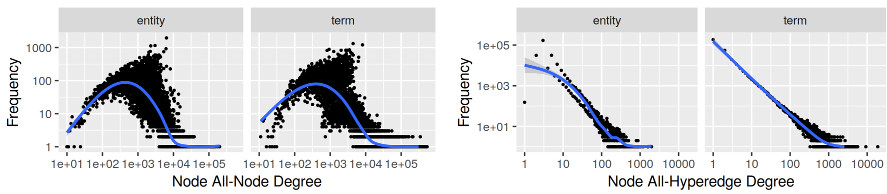
Global Statistics In Table 1, we present several global statistics about the hypergraph-of-entity, in particular the number of nodes and hyperedges, discriminated by type, the average degree, the average clustering coefficient, the average path length, the diameter and the density. The average clustering coefficient was computed based on a sample of 5,000 nodes and a sample of 100,000 neighbors for each of those nodes. The average path length and the diameter were computed based on a sample of shortest distances between 30 random pairs of nodes and

Table 1: Global statistics

Statistic	**Value**
Nodes	607,213
term	323,672
entity	283,541

Statistic	**Value**
Hyperedges	253,154
Undirected	14,938
document	7,484
related_to	7,454
Directed	238,216
contained_in	238,216

Statistic	**Value**
Avg. Degree	0.8338
Avg. Clustering Coefficient	0.1148
Avg. Path Length	8.3667
Diameter	17
Density	3.88e-06



(a) Based on connected nodes. (b) Based on connected hyperedges.

Fig. 1: Node degree distributions (log-log scale).

the intersections of 1,000 random walks of length 1,000 launched from each element of the pair. Finally, the density was computed based on Equation 1. As we can see, for the 7,487 documents the hypergraph contains 607,213 nodes and 253,154 hyperedges of different types, an average degree lower than one (0.83) and a low clustering coefficient (0.11). It is also extremely sparse, with a density of $3.9e{-}06$. Its diameter is 17 and its average path length is 8.4, almost double when compared to a social network like Facebook [2].

Local Statistics Figure 1 illustrates the node degree distributions. In Figure 1a, the node degree is based on the number of connected nodes, with the distribution approximating a log-normal behavior. In Figure 1b, the node degree is based on the number of connected hyperedges, with the distribution approximating a power law. This shows the usefulness of considering both of the node degrees in the hypergraph-of-entity, as they are able to provide different information.

Figure 2 illustrates the hyperedge cardinality distribution. For *document* hyperedges, cardinality is log-normally distributed, while for *related_to* hyperedges the behavior is slightly different, with low cardinalities having a higher frequency than they would in a log-normal distribution. Finally, the cardinality distribution of *contained_in* hyperedges, while still heavy-tailed, presents an initial linear behavior, followed by a power law behavior. The maximum cardinality for this type of hyperedge is also 16, which is a lot lower when compared to *document* hyperedges and *related_to* hyperedges, which have cardinality 8,167 and 3,084, respectively. This is explained by the fact that *contained_in* hyperedges establish a directed connection between a set of terms and an entity that contains those terms, being limited by the maximum number of words in an entity.

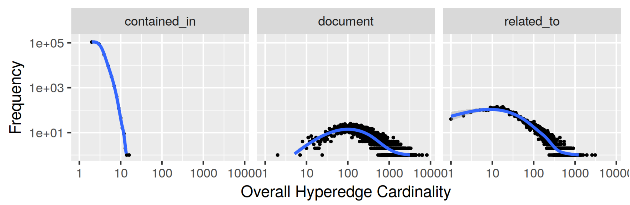
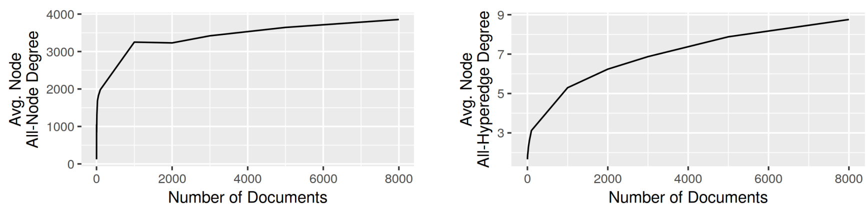


Fig. 2: Hyperedge cardinality based on the total number of nodes (log-log scale).



(a) Based on connected nodes. (b) Based on connected hyperedges.

Fig. 3: Average node degree over time.

Temporal Statistics In order to compute temporal statistics, we first generated 14 snapshots of the index based on a limit L of documents, for $L \in \{1, 2, 3, 4, 5, 10, 25, 50, 100, 1000, 2000, 3000, 5000, 8000\}$.

Figure 3 illustrates the node-based and hyperedge-based average node degrees over time (represented as the number of documents in the index at a given instant). As we can see, both functions tend to converge, however this is clearer for the node-based degree, reaching nearly 4,000 nodes, through only 9 hyperedges, on average. Figure 4 illustrates the average undirected hyperedge cardinality over time, with a convergence behavior that approximates 300 nodes per hyperedge, after rising to an average of 411.88 nodes for $L = 25$ documents.

Figure 5 illustrates the evolution of the average path length and the diameter of the hypergraph over time. For a single document, these values reached 126.1 and 491, respectively, while, for just two documents, they immediately lowered to 3.8 and 10. For higher values of L, both statistics increased slightly, reaching 7.2 and 15 for the maximum number of documents. Notice that these last values are equivalent to those computed in Table 1 (8.4 and 17, respectively), despite resulting in different quantities. This is due to the precision errors in our estimation approach, resulting in a difference of 1.2 and 2, respectively, which is tolerable when computation resources are limited. In Figure 6, we illustrate the evolution of the clustering coefficient, which rapidly decreases from 0.59 to 0.11. The low average path length and clustering coefficient point towards a weak community structure, possibly due to the coverage of divergent topics. However, we would require a random generation model for hypergraphs, like the Watts–Strogatz model for graphs, in order to properly interpret the statistics.

Figure 7 illustrates the evolution of the density over time. The density is consistently low, starting from 1.37e−03 and progressively decreasing to 3.91e−06

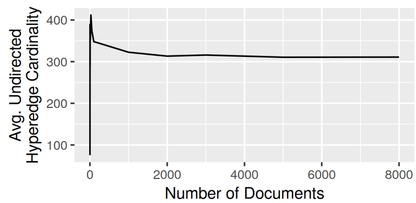


Fig. 4: Average hyperedge cardinality over time.

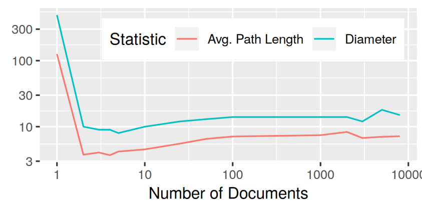


Fig. 5: Average estimated diameter and average shortest path over time.

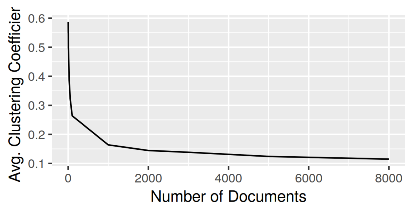


Fig. 6: Average estimated clustering coefficient over time.

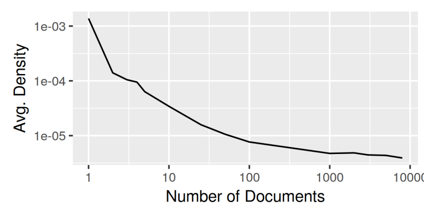


Fig. 7: Average density over time.

as the number of documents increases. This shows that the hypergraph-of-entity is an extremely sparse representation, with limited connectivity, which might benefit precision in a retrieval task.

Finally, we also measured the space usage of the hypergraph, both in disk and in memory. In disk, the smallest snapshot required 43.8 KiB for one document, while the largest snapshot required 181.9 MiB for the whole subset. Average disk space over all snapshots was 37.5 MiB ± 58.9 MiB. In memory, the smallest snapshot used 1.0 GiB for one document, including the overhead of the data structures, and the largest snapshot used 2.3 GiB for the whole subset. Average memory space over all snapshots was 1.3 GiB ± 461.1 MiB. Memory also grew faster for the first 1,000 documents, apparently leading to an expected convergence, although we could not observe it for such a small subset.

5 Conclusions

We have characterized the hypergraph-of-entity representation model, based on the structural properties of the hypergraph. We analyzed the node degree distributions, based on nodes and hyperedges, and the hyperedge cardinality distributions, illustrating their distinctive behavior. We also analyzed the temporal behavior, as documents were added to the index, studying average node degree and hyperedge cardinality, estimated average path length, diameter and clustering coefficient, as well as density and space usage requirements. Our contributions go beyond the characterization of the hypergraph-of-entity, as we show an application of two approximation approaches for computing statistics based on the shortest distance, as well as the clustering coefficient. We also proposed a simple approach for computing the density of a general mixed hypergraph, based on the corresponding bipartite mixed graph.

In the future, we would like to further explore the computation of density, as the bipartite-based density we proposed, although useful, only accounts for hyperedges already in the hypergraph. We would also like to study the parameterization of the two estimation approaches we proposed, based on random walks and node sampling. Another open challenge in hypergraphs is the definition of random generation model, which would be useful to improve characterization. Finally, several opportunities also exist in the study of the hypergraph at a mesoscale, be it identifying communities, network motifs or graphlet, or exploring unique patterns to hypergraphs.

Acknowledgements

José Devezas is supported by research grant PD/BD/128160/2016, provided by the Portuguese national funding agency for science, research and technology, Fundação para a Ciência e a Tecnologia (FCT), within the scope of Operational Program Human Capital (POCH), supported by the European Social Fund and by national funds from MCTES.

References

1. Aparicio, D., Ribeiro, P., Silva, F.: Graphlet-orbit transitions (got): A fingerprint for temporal network comparison. PLoS One 13, e0205497 (October 2018)
2. Backstrom, L., Boldi, P., Rosa, M., Ugander, J., Vigna, S.: Four degrees of separation. CoRR abs/1111.4570 (2011), http://arxiv.org/abs/1111.4570
3. Bast, H., Buchhold, B.: An Index for Efficient Semantic Full-text Search. In: Proceedings of the 22Nd ACM International Conference on Conference on Information and Knowledge Management. pp. 369–378 (2013), http://doi.acm.org/10.1145/2505515.2505689
4. Bast, H., Buchhold, B., Haussmann, E., et al.: Semantic search on text and knowledge bases. Foundations and Trends® in Information Retrieval 10(2-3), 119–271 (2016)
5. Bastian, M., Heymann, S., Jacomy, M.: Gephi: An open source software for exploring and manipulating networks. In: Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17-20, 2009 (2009), http://aaai.org/ocs/index.php/ICWSM/09/paper/view/154
6. Berge, C.: Graphes et hypergraphes. Dunod: Paris (1970)
7. Bhagdev, R., Chapman, S., Ciravegna, F., Lanfranchi, V., Petrelli, D.: Hybrid search: Effectively combining keywords and semantic searches. In: European Semantic Web Conference. pp. 554–568. Springer (2008)
8. Brandes, U., Eiglsperger, M., Herman, I., Himsolt, M., Marshall, M.S.: Graphml progress report structural layer proposal. In: International Symposium on Graph Drawing. pp. 501–512. Springer (2001)

9. Csardi, G., Nepusz, T., et al.: The igraph software package for complex network research. InterJournal, Complex Systems 1695(5), 1–9 (2006)
10. Devezas, J., Nunes, S.: Hypergraph-of-entity: A unified representation model for the retrieval of text and knowledge. Open Computer Science 9(1), 103–127 (Jun 2019), https://doi.org/10.1515/comp-2019-0006
11. Estrada, E., Rodriguez-Velazquez, J.A.: Complex networks as hypergraphs. arXiv preprint physics/0505137 (2005)
12. Fernández, J.D., Martínez-Prieto, M.A., de la Fuente Redondo, P., Gutiérrez, C.: Characterizing rdf datasets. Journal of Information Science 1, 1–27 (2016)
13. Gallagher, S.R., Goldberg, D.S.: Clustering coefficients in protein interaction hypernetworks. In: ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics. ACM-BCB 2013, Washington, DC, USA, September 22-25, 2013. p. 552 (2013), https://doi.org/10.1145/2506583.2506635
14. Ge, W., Chen, J., Hu, W., Qu, Y.: Object link structure in the semantic web. In: The Semantic Web: Research and Applications, 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 3, 2010, Proceedings, Part II. pp. 257–271 (2010), https://doi.org/10.1007/978-3-642-13489-0_18
15. Głąbowski, M., Musznicki, B., Nowak, P., Zwierzykowski, P.: Shortest path problem solving based on ant colony optimization metaheuristic. Image Processing & Communications 17(1-2), 7–17 (2012)
16. Halpin, H.: A query-driven characterization of linked data. In: Proceedings of the WWW2009 Workshop on Linked Data on the Web, LDOW 2009, Madrid, Spain, April 20, 2009. (2009), http://ceur-ws.org/Vol-538/ldow2009_paper16.pdf
17. Himsolt, M.: GML: A portable graph file format. Tech. rep., Universität Passau (1997)
18. Klamt, S., Haus, U., Theis, F.J.: Hypergraphs and cellular networks. PLoS Computational Biology 5(5) (2009), https://doi.org/10.1371/journal.pcbi.1000385
19. Li, D.: Shortest paths through a reinforced random walk. Tech. rep., University of Uppsala (2011)
20. Mubayi, D., Zhao, Y.: Co-degree density of hypergraphs. J. Comb. Theory, Ser. A 114(6), 1118–1132 (2007), https://doi.org/10.1016/j.jcta.2006.11.006
21. Ouvrard, X., Goff, J.L., Marchand-Maillet, S.: Adjacency and tensor representation in general hypergraphs part 1: e-adjacency tensor uniformisation using homogeneous polynomials. CoRR abs/1712.08189 (2017), http://arxiv.org/abs/1712.08189
22. Ribeiro, B.F., Basu, P., Towsley, D.: Multiple random walks to uncover short paths in power law networks. In: 2012 Proceedings IEEE INFOCOM Workshops, Orlando, FL, USA, March 25-30, 2012. pp. 250–255 (2012), https://doi.org/10.1109/INFCOMW.2012.6193500
23. Voorhees, E.M.: The efficiency of inverted index and cluster searches. In: SIGIR'86, Proceedings of the 9th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Pisa, Italy, September 8-10, 1986. pp. 164–174 (1986), https://doi.org/10.1145/253168.253203
24. Yu, W., Sun, N.: Establishment and analysis of the supernetwork model for nanjing metro transportation system. Complexity 2018, 4860531:1–4860531:11 (2018), https://doi.org/10.1155/2018/4860531
25. Zobel, J., Moffat, A., Ramamohanarao, K.: Inverted files versus signature files for text indexing. ACM Trans. Database Syst. 23(4), 453–490 (1998), http://doi.acm.org/10.1145/296854.277632