

TDK dolgozat

Szerző: Ficsor Attila

Frequent itemset mining can measure how global events
thematize the media

Szerző: Ficsor Attila

Adattudomány MSc. 3. félév

Témavezetők: Dr. Sebestyén Viktor tudományos főmunkatárs, Cvetkó Tímea PHD hallgató

HUN-REN-PE Complex Systems Monitoring Research Group

Frequent itemset mining can measure how global events thematize the media

*Készítette: Ficsor Attila
Adattudomány MSc.*

HUN-REN-PE Complex Systems Monitoring Research Group

Témavezető: Dr. Sebestyén Viktor, Czvetkó Tímea

The emerging complexity of societal problems requires international focus and cooperation in finding solutions. International events serve the purpose of addressing challenges and raising public awareness. The impact of events on public awareness can be assessed by analyzing world news, as these news are linked to taxonomies, allowing for identifying and categorizing news topics. The impact of an event can be detected by analyzing temporal changes in frequent news topic patterns, which allows for the identification of how events shape media topics and, therefore, public sensitivity.

Considering the complex interplay between media topics is essential to analyze frequent itemsets within a context. Therefore, this study provides a methodology for identifying the effect of events using frequent itemset mining of world news and its visualization and analysis through hypergraphs and its temporal changes. Hypergraphs allow us to analyze and visualize the group interactions between news topics and the frequent itemsets of news topics. By comparing hypergraphs over time, the shifts of event topics can be detected, providing additional information about temporal focus areas. Additionally, information values of the frequent patterns are applied to measure a given pattern's surprising effect.

Since the existing hypergraph visualization tools have limitations regarding customisability, representing the specific analyzations and visualizations discussed in this work proved to be quite difficult. To counter the limitations of previous tools, to enhance the visualization of the results, and make the understanding of the frequent patterns easier, a novel open-source hypergraph visualization tool was developed using the Godot engine. This tool is fully customizable, allowing users to manually modify the visualization of the hypergraph, and even compare multiple hypergraphs on the same visualization.

The applicability of the methodology is demonstrated through the analysis of the Conference of Parties (COP) international sustainability events.

Contents

Nomenclature	4
1 Introduction	6
2 Methodology of hypergraphs-based frequent news itemset analysis	10
2.1 Data acquisition	10
2.2 Aggregation of news data	11
2.3 Frequent itemset mining of world news	11
2.4 Information value analysis of frequent itemsets	12
2.5 Hypergraph representation of frequent itemsets	13
2.6 Summary of the methodology	15
3 Hypergraph Visualization Tool Built with the Godot Engine	18
3.1 Software description	19
4 Results	23
5 Conclusion	32
6 Future plans	33
Bibliography	39
A Appendix	40
A.1 Example data accepted by HyperGodot	40
A.2 Andrew’s monotone chain GDScript implementation	40

Nomenclature

$\text{SumInf}(F)$ Overall information value of all patterns in F

$C(v_i)$ Degree centrality of vertex v_i

D Dataset comprising multiple itemsets T_k

D_{ts}^{ts+w} Dataset of itemsets within the time window $[ts, ts + w]$

E Family of hyperedges in the hypergraph

F Set of all frequent itemsets that reach a specified threshold

F_i Frequent pattern, a subset of items I

$H = (V, E, W)$ Incidence matrix of the hypergraph

I Set of all possible items i_j

i_j An object extracted from articles (e.g., location, theme, person, organization)

$\text{Inf}(F_i)$ Information value of pattern F_i

$M(F_i)$ Set of supersets of F_i that are frequent

$P(F_i)$ Probability of encountering the frequent pattern F_i

$\text{supp}(F_i)$ Support of itemset F_i in the full dataset

$\text{supp}_{ts}^{ts+w}(F_i)$ Support of itemset F_i within a specific time window $[ts, ts + w]$

threshold Support threshold for frequent itemsets

T	Represents an itemset of objects extracted from news
T_k	Itemset with a time label, where $T_k = (\{i_1, i_2, \dots, i_j\}, t_k)$
ts	Start of a time window
w	Duration of a time window
V	Set of vertices in the hypergraph
W	Weights associated with each edge in the hypergraph

1. Chapter

Introduction

The convergence of interconnected crises, or 'polycrises,' creates complex global challenges that demand innovative tools for analysis and understanding. The University of Pannonia is committed to sustainability solutions. I have the opportunity to work within the framework of Multidisciplinary National Laboratory for Climate Change (RRF-2.3.1-21-2022-00014), which is dedicated to advancing sustainability research and solutions at the University of Pannonia. Thereby, my research is aligned with the scope of providing novel tools and methods to understand sustainability-related issues.

International events on global issues aim to foster the world's efforts to address global challenges and raise public awareness and sensitivity towards issues such as climate change, poverty, equality, and justice among others [1]. The importance of these events has been increased due to the emerging complexity of global challenges and accelerated actions are required across all areas [2]. Therefore, there is a need to measure the effect of these events and their impact on public awareness. Events not only reflect the collective sentiment of people, but also articulate community values, visions, and aspirations [3]. A possible approach for identifying public awareness of global issues is news analysis [4]. The interaction between news topics must be explored to analyze the impact of the events in context. News analysis can provide a tool for measuring and analyzing global news occurrences over time. For which, this study presents a tool that enables us to detect the effect of an event by analysing the news over time. The methodology is based on the temporal hypergraph-based representation of frequent news itemsets and its qualification is based on information value analysis to identify how global events thematize the media.

The world news is gathered from The GDELT Project (Global Database of

Events, Language and Tone) which is an open platform monitoring the news media of the world in over 100 languages and identifies people, locations, organizations, themes, sources, emotions, counts, quotes, images and events related to the news. Its vast dataset, encompassing diverse sources and languages, facilitates a nuanced understanding of events and their implications. The news has been linked to taxonomies such as the World Bank Topical Taxonomy [5], or CrisisLex [6], thereby the news topics are categorized.

Prior research has underscored the importance of news analysis as it has been utilized in for example understanding environmental change through social data gathered from the news [7], forecasting hospitalization rate [8], detecting cyber security issues [9], or in understanding investor reaction to sustainability news in energy companies [10]. News can also be used to proxy the peace index of countries which is a principal dimension of well-being [11]. On sustainability issues, identifying the interconnectedness of Sustainable Development Goals has also been explored through the network analysis of world news [12], however, the study is limited to a pairwise network representation of sustainability events for a specific time frame, leaving the cooccurrence of news domains and the exploration of potential sequences of sustainability issues unexplored. Considering the spatiotemporal characteristics of the events such as COVID-19 their influencing factors can be identified [13], media spotlights analyzed [14] and temporal itemsets in complex event data can be identified [15]. Furthermore, the spatial and temporal dynamic graph representation of news events have been made to predict the occurrence of future social events for the target country [16]. The network-based representation of news data and their analysis has been utilized e.g. for evaluating and improving social awareness of energy communities [17], detecting and analyzing fake news [18], or analyzing news appearing in various countries on a specific event through semantic network analysis [19]. Beyond the pairwise connections, a hypergraph-based assessment has also been made to explore the impact of financial news on market behaviors by geometric hypergraph attention network (GHAN) [20]. The hypergraph-based assessment can allow e.g. cross-context collaborative associations to support context-aware prediction [21], however, it must be noted that the utilization of hypergraph-based representation of world news has not been studied extensively. For analyzing GDELT-based news domains that are

frequently observed together and applying association rule mining for identifying possible relations between the domains has been proven to be efficient [22], however, the study only applies this approach for identifying co-location itemsets, neglecting the broader potential for uncovering intricate interdependencies among sustainability issues and news domains such as events, people, countries. Frequent itemset mining has been also utilized in fake news discovery [23], and knowledge discovery from news events [24].

While previous studies have applied network-based representations and frequent itemset mining to analyze news interactions around specific themes or events, most research has been limited to pairwise connections, co-location itemsets, or analyses confined to specific event types or domains. There has been limited exploration of how global events dynamically influence and shift the media landscape over time through complex, multi-domain interactions represented in a hypergraph structure. Furthermore, the use of temporal hypergraphs, particularly for capturing the evolution of news thematization around sustainability and societal issues, remains underexplored. This gap presents an opportunity to investigate how hypergraph-based methods can identify nuanced, temporal itemsets of public awareness and sentiment in response to major global events across diverse news topics.

Therefore this study presents a methodology utilizing frequent itemset mining to identify how news topics are grouped over time near global events thereby, measuring how global events thematize the media. frequent news itemsets are discovered at diverse time slots and compared over time to see the changes in media focus. The frequent news itemsets represented as hypergraphs allow the interpretation of group interactions of news topics and identifying temporal focus areas. For the qualification of frequent itemsets, information value analysis was used indicating the diversity of news topics.

The temporal hypergraph-based representation of frequent news itemsets and the qualification of the effect of global news based on information value analysis has not been studied extensively, however, it holds great potential in identifying the effect of certain events and how they appear in public and news focus/news thematization change over time. It can be utilized to measure how global events appear in public media, thereby shaping public awareness of societal issues.

The analysis detailed in this study requires a visualization tool that allows for manual modification, temporal hypergraph representation, and high customization. For these reasons a novel hypergraph visualization tool was developed, to suit the specific needs of this project. The HyperGodot tool helps with understanding the frequent itemsets, and by comparing different hypergraphs, the differences can be noted and analyzed further.

The structure and contribution of the study are as follows.

- First, the hypergraph-based frequent news itemset analysis methodology is introduced and its qualification through information value analysis.
- Hypergraph visualization tool was developed to suit the study's specific needs.
- The method is then applied to a case study on identifying news thematization around the Conference of Parties (COP) events held in 2021, 2022, and 2023.

The results underline the method's applicability as the shifts in media focus are well visible over the years. As well as the conference-related topics and frequent itemsets show a significant increase in media attention, but not before and after the conference.

2. Chapter

Methodology of hypergraphs-based frequent news itemset analysis

This chapter outlines the methodology for analyzing frequent news itemsets and using hypergraph-based approaches, offering a structured framework to explore complex associations within media coverage. First, the the process of data acquisition, the aggregating news data, mining frequent itemsets, and related metrics is formalized, followed by a detailed description of the hypergraph-based visualization and analysis of the frequent itemsets and their temporal changes. Finally, a summary of the methodology is provided, outlining the steps to analyze frequent itemsets to define how global events thematize the media.

2.1 Data acquisiton

The news data is derived from the GDELT Global Knowledge Graph [25] which includes several objects of interest such as themes, regions, people, and organizations. These components are crucial for describing the main aspects of a news story. By using this information, world events can be better filtered and analyzed, helping to understand and interpret global news more effectively. The GDELT Database has been linked to taxonomies such as the World Bank Topical Taxonomy [5] or the CrisisLex [6], which enables the definition of the topics of interest (for example, the theme of an event) based on the taxonomy models, which will serve as the categorization of the derived news. Topics must be selected in the taxonomic model to retrieve topic-relevant news data. For example, the aim is to derive news data regarding climate change-related events and identify related topics from e.g. World Bank Topical Taxonomy, non-

exhaustively e.g. Climate Change Mitigation, Climate Resilient Development etc.

2.2 Aggregation of news data

Each news article is considered an itemset with a time label, $T_k = (\{i_1, i_2, \dots, i_j\}, t_k)$, where T is representing an itemset of objects extracted from the news, used as a transaction, and k represents a timestamp. An item i_j represents an object extracted from the articles, such as a location, theme, person, or organization. The set I contains all possible items i_j . The dataset $D = \{T_1, T_2, \dots, T_n\}$ for a s comprises multiple itemsets T_k . To analyze data over a specific time window $[ts, ts + w]$, where w represents the duration of the time window (e.g., 24 hours), the dataset is defined as $D_{ts}^{ts+w} = \{T_k \mid ts < t_k < ts + w\}$ which contains all itemsets from ts to $ts + w$.

2.3 Frequent itemset mining of world news

A frequent itemset $F_i \subseteq I$ is an itemset that appears frequently within the dataset. The support of an itemset F_i within the full timeframe is defined as:

$$\text{supp}(F_i) = \frac{|\{T_k \mid F_i \subseteq T_k\}|}{|D|} \quad (2.1)$$

where $|D|$ is the total number of transactions in the full dataset.

The support of an itemset F_i within a specific time window $[ts, ts + w]$ is defined as:

$$\text{supp}_{ts}^{ts+w}(F_i) = \frac{|\{T_k \mid F_i \subseteq T_k \text{ and } ts < t_k < ts + w\}|}{|D_{ts}^{ts+w}|} \quad (2.2)$$

where $|D_{ts}^{ts+w}|$ is the total number of transactions within the time window.

Frequent itemsets F is all the frequent itemsets that reach a specified threshold value:

$$F_{ts}^{ts+w} = \{F_i \mid \text{supp}_{ts}^{ts+w}(F_i) > \text{threshold}\} \quad (2.3)$$

A maximal frequent itemset is an itemset that is frequent and has no superset

that is also frequent. Itemset $F_i \subseteq I$ is considered maximal if and only if:

$$M(F_i) = \{F_j \subseteq I \mid F_i \subset F_j \text{ and } \text{supp}(F_j) > \text{threshold}\} = \emptyset \quad (2.4)$$

where $M(F_i)$ is the set of all supersets F_j of F_i that are frequent. A maximal frequent itemset F_i is one for which no proper superset is frequent, indicating that it is the largest frequent itemset within the transaction database.

2.4 Information value analysis of frequent itemsets

Calculating the information value can help in understanding key trends in the news data. The information value, derived from these probabilities of the given frequent itemsets appearing within all the frequent itemsets for a given timeframe, provides insight into the diversity of the topics present within the itemsets.

The probability $P(F_i)$ of encountering a specific frequent itemset F_i is given by:

$$P(F_i) = \text{supp}(F_i) \quad (2.5)$$

The information value of a given itemset can be calculated by:

$$\text{Inf} = \log_2 \frac{1}{P(F_i)} \quad (2.6)$$

The overall information value of all itemsets in F can be further computed, to showcase the changes in the data over time. This is done by summing the information values of each itemset $|F|$:

$$\text{SumInf}(F) = \sum_{i=1}^{|F|} \text{Inf}(F_i) \quad (2.7)$$

By repeating this average calculation for multiple timeframes, a time series of the average information values is obtained, revealing how the frequent itemsets' diversity evolves.

2.5 Hypergraph representation of frequent itemsets

Frequent itemsets can be represented as a hypergraph [26], as an invaluable tool to explore the multilevel connections between the objects that were defined from the news.

A hypergraph is a generalization of a graph in which an edge can join any number of vertices, in contrast to an ordinary graph, where an edge connects exactly two vertices. A hypergraph can be represented by the incidence matrix $H = (V, E, W)$. In this structure, $V = \{v_i\}_{i=1}^n = \{i_j \in I \mid \exists F_i \in F_{ts}^{ts+w}, i_j \in F_i\}$ is a set of vertices formed by all the elements represented in all of the frequent itemsets inside the given timeframe, $E = \{e_j\}_{j=1}^m = F_i \in F_{ts}^{ts+w}$ indicates a family of hyperedges where each e_j is a subset of V , and equals to a single frequent itemset F_i , and $W = \text{supp}_{ts}^{ts+w}(F_i)$ represents the weights associated with each edge, which are the support values of each frequent itemset forming the given edge. [27]. The incidence matrix H defines the relation between vertices and hyperedges, with $H(v_i, e_j) = 1$ if the item v_i is part of the itemset e_j , and $H(v_i, e_j) = 0$ otherwise.

By considering two hypergraphs, F_{ts}^{ts+w} and F_{ts+w}^{ts+2w} , representing frequent itemsets over consecutive time windows, temporal changes can be effectively analyzed in the data. Naturally, nonconsecutive hypergraphs can be analyzed in this manner too but in the scope of this study, the aim is to compare consecutive timeframes. To achieve this, a merged or temporal hypergraph is constructed, that captures the differences between these two timeframes. Specifically, a hypergraph is generated from both sets of frequent itemsets and the common elements shared between the two timeframes are removed. Both hypergraphs are visualized on a single merged hypergraph by distinguishing the edges corresponding to each timeframe. Edges from the first hypergraph (F_{ts}^{ts+w}) are represented with solid lines, while those from the second hypergraph (F_{ts+w}^{ts+2w}) are depicted with segmented lines. Degree centrality is used to highlight the most important themes. Degree centrality quantifies the significance of a vertex by measuring the number of hyperedges connected to it, showing which topics are most central within the hypergraph. Vertices with a higher degree centrality represent themes that are connected to more topics, indicating their higher importance within the dataset. The degree centrality

$C(v_i)$ for a vertex v_i is calculated by:

$$C(v_i) = \sum_{j=1}^m H(v_i, e_j) \quad (2.8)$$

An example incidence matrix of a merged temporal hypergraph H that contains 8 vertices, 6 hyperedges and incorporates data from 2 consecutive timeframes, F_{ts}^{ts+w} (2022) and F_{ts+w}^{ts+2w} (2023), is provided on Table 2.1, and visualized in Figure 2.1, where the hyperedge width is proportional with the weight (support value) of each hyperedge.

Vertex	Label
v_1	Environment and Natural Resources
v_2	Pollution Management
v_3	Energy and Extractives
v_4	Energy Efficiency
v_5	Health Emergency Preparedness and Disaster Response
v_6	Air Pollution
v_7	Climate Change
v_8	Renewable Energy

$$H = \begin{matrix} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & \text{Weight} & \text{Group} \\ \begin{matrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{matrix} & \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} & \begin{matrix} 2662 \\ 3163 \\ 7720 \\ 3312 \\ 3301 \end{matrix} & \begin{matrix} 2022 \\ 2022 \\ 2022 \\ 2023 \\ 2023 \end{matrix} \end{matrix}$$

Table 2.1: Incidence matrix of example hypergraph H

In Figure 2.1 hypergraph H is represented, where the hyperedge thickness is proportional to the hyperedge weight (support value), solid lines represent the first timeframe, and segmented lines represent the second timeframe and node size corresponds to the degree centrality. v_1 - Environment and Natural Resources, and v_3 Energy and Extractives have the highest degree centrality. It is well shown that compared to 2022, in 2023 new frequent topics appeared such as the itemset of "Energy and Natural Resources - Energy and Extrac-

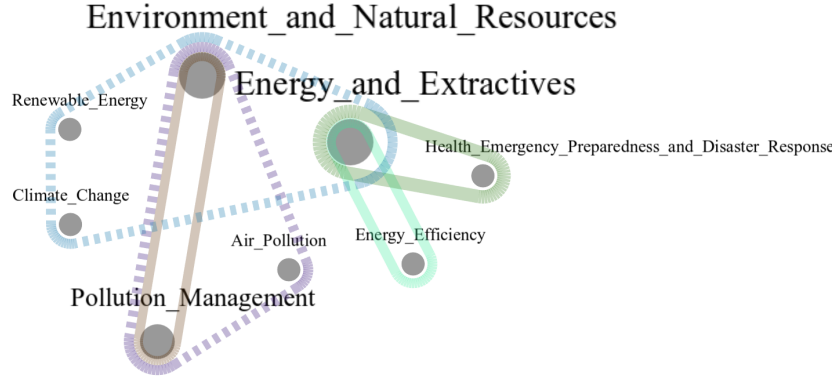


Figure 2.1: Visualization of example hypergraph H , where the hyperedge thickness is proportional to the hyperedge weight (support value), solid lines represent the first timeframe, and segmented lines represent the second timeframe and node size corresponds to the degree centrality.

tives - Climate Change - Renewable Energy", alongside "Energy and Natural Resources - Air Pollution - Pollution Management". Therefore the shift in topical focus can be seen. The shift toward air pollution-related actions can be due to wildfires that occurred in Australia and Canada in 2023, which caught the attention of the media.

2.6 Summary of the methodology

The summary of the methodological steps of the proposed hypergraph-based frequent news itemset analysis are indicated in Figure 2.2.

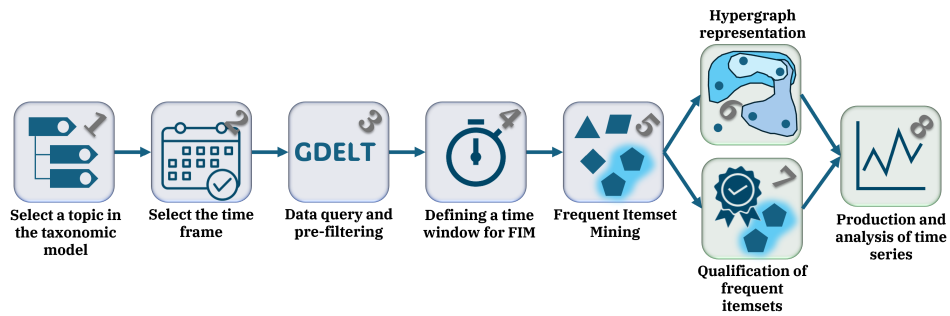


Figure 2.2: Workflow for identifying the media thematizing effect of global events.

The steps are the following:

1. Select a topic in taxonomic model:

Relevant topics are identified from taxonomies such as the World Bank Topical Taxonomy or CrisisLex, which categorize news data based on themes of interest. These objects form the set of I that contains all items i_j , which are in this case news topics appearing in D .

2. Select the time frame:

The selection of the timeframe (from ts to $ts + w$, where w is a specific time window) is up to the user. For example, in the presented case, the whole time frame is a month before the event until the month after the event desired to be explored. For example, the 26th Conference of Parties event was held in Glasgow from 2021-10-31 until 2021-11-12, the selected timeframe was $ts = 2021-10-01$ until $ts + w = 2021-10-01 + 72 \text{ days} = 2021-12-12$.

3. Data query and pre-filtering:

Data is queried from the GDELT Global Knowledge Graph [25] for the defined topics and selected timeframe. Data is queried through Google BigQuery for the selected topics and time frame. In this paper, the database is filtered to only the themes, however, the analysis can be made to reveal further information e.g. regarding the tones and locations.

4. Defining a time window for frequent itemset mining:

The set of time window (w) for frequent itemset mining is up to the user. It will serve as a unit where the frequent itemset mining is derived. For example, in the case of this study the $w = 1$ day, thereby frequent itemsets are derived for each day within the whole time frame.

5. Frequent itemset mining:

Frequent itemset F_i and support values are defined based on Equations 2.1-2.4. It makes detecting e.g. overall, or daily changes in topical focus (frequently coocurred themes) in the media possible.

6. Qualification of frequent itemsets: Frequent itemsets F_i can be qualified through information value Inf , indicated in Equations 2.5 - 2.7.

Information value measures the e surprising effect of the itemsets.

7. Hypergraph representation of frequent itemsets:

Frequent itemsets F_i can be represented as hypergraphs $H = (V, E, W)$, revealing cross-interactions within the network. the whole frequent itemset can be visualized, or any set operation of frequent itemsets can be represented, such as:

- (a) Union: $F_i \cup F_j$, includes all items from both frequent itemsets. It can represent the overall thematic coverage of e.g. conferences over years.
- (b) Intersection: $F_i \cap F_j$, includes only the frequent itemsets represented in both frequent itemsets. This can represent the common themes that appear in e.g. each year at conferences.
- (c) Difference: $F_i \setminus F_j$, includes the itemsets that are only in one of the frequent itemsets. This reveal thematic appearances that are only characterise e.g. one of the conferences.

Furthermore, network metrics can be utilized and used in visualization, such as degree centrality (Equations 2.8) and edge weights (support value).

8. Production and analysis of time series:

It involves the analysis of frequent itemsets over time using support metrics, itemset counts, and information values to uncover temporal trends and dynamics. By observing how these metrics evolve, insights can be gained into the persistence, fluctuation, or emergence of specific itemsets, helping to contextualize their importance and identify shifts in relevance over time.

3. Chapter

Hypergraph Visualization Tool Built with the Godot Engine

Existing tools for hypergraph visualization offer robust analysis capabilities. In Table 3.1, available hypergraph visualization libraries are gathered and features are interpreted. It is clearly visible, that most of the libraries often lack in-depth customization. Specifically, when displaying maximal frequent item-sets, nodes, and edges must be well-positioned to avoid overlap and enhance clarity in the hypergraph visualization. Many existing solutions fall short in this regard, as they offer limited options for manual node placement and property adjustments. For example, HyperNetX includes an interactive widget that allows basic exploration of the hypergraph, but its functionality remains quite restricted for more complex visual adjustments. Furthermore, features to distinguish hyperedges by class or by choice are often neglected in the available tools. Therefore, there is a need for a targeted hypergraph visualization tool that allows in-depth customization features. I present a novel tool, called HyperGodot, which overcomes the above-mentioned limitations.

The hypergraph visualization tool HyperGodot was developed to address the limitations of existing hypergraph visualization software, which often lacks the flexibility required for specific analysis and manipulation. This tool allows users to manually position nodes, adjust node and edge colors, toggle the visibility of specific edges, compare multiple hypergraphs as a merged temporal hypergraph, toggle the centrality value-based node size, etc.

Tool	Language	Feature	Reference
HAT: Hypergraph Analysis Toolbox	MATLAB, Python	construction, visualization, expansion, similarity measures, centrality, entropy, tensor-based analysis, controllability	[28]
HyGrCIs	MATLAB	clustering	[29]
Hypergraph edge/vertex matrix	MATLAB	construction	[30]
HyperNetX	Python	clustering, visualization, homology, clustering	[31]
Multihypergraph	Python	creation of looped-multi-hyper-graphs	[32]
Hypergraph-Vis	Python	construction, visualization, simplification	[33]
Hypergraphx	Python	construction, visualization, centrality, statistics, analysis, filters, dynamics	[34]
HALP: Hypergraph Algorithms Package	Python	directed hypergraphs, walks, partitioning	[35]
HyperG	R	hypergraph creation, clustering, graph based representations and calculations, visualization	[36]
hMETIS	C/C++	partitioning, implemented in parallel	[37]
Phoenix	C/C++	clustering, implemented in parallel	[38]
HyperGraphLib	C/C++		[39]
HyperGraphs.jl	Julia	construction, calculate properties	[40]
Chapel Hypergraph Library (CHGL)	Chapel	construction, calculate properties	[41]
HyperGodot	GDScript	construction, visualization, centrality, manual visualization modification temporal hypergraph visualization	

Table 3.1: Available hypergraph libraries

3.1 Software description

HyperGodot is developed using the Godot engine 4.3 version [42], using the engine’s own scripting language GDScript. Godot is an open-source video game engine, used widely in indie videogame development, simulations, and any real-time-sensitive projects. Its growing popularity is explained by its compact size, easy-to-learn scripting language, and ever-growing user base. For this project, a real-time engine was needed, so we can manually adjust the hypergraph, so Godot was a natural choice.

The biggest difference between HyperGodot and the available libraries is the ability to manually modify the visualization during runtime, including repositioning the nodes, changing the color of the edges, changing the function calculating the edge width, and visualizing different groups, which makes this software able to visualize temporal hypergraphs. This ability makes it possible to compare multiple hypergraphs of a different timestamps, or simply compare 2 hypergraphs based on the same or overlapping nodes.

The summary of the methodology can be seen on Figure 3.1:

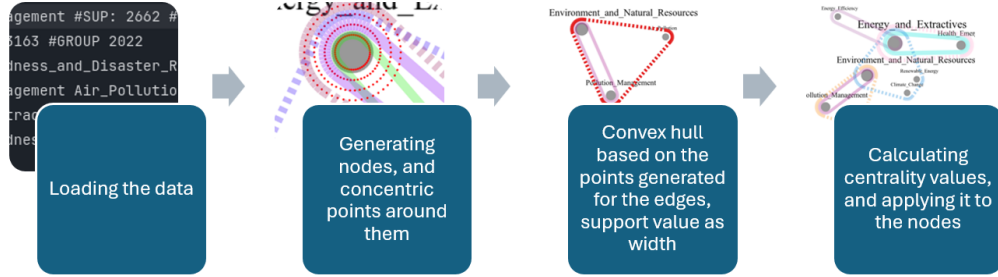


Figure 3.1: Summary of the methodology employed in the hypergraph visualization tool

1. **Loading of the data:** The data structure accepted is based on the SPMF software's output, as it is used in the methodology to mine frequent patterns. The output text is expanded with the `#GROUP` tag, which contains the data differentiating the distinct hypergraphs we want to compare. Example of the accepted data can be seen in the Appendix A.1
2. **Generating nodes and concentric points:** Each unique entry before the `#SUP` tag is going to be a unique node. The SPMF output indices are translated to the real name of the given themes, so it is more readable. The tool's methodology generates circles around each node based on the number of edges connected to it. Each edge generates a circle around all nodes it links, and points are generated along each circle to enable convex hull calculations. The number of points on each circle is adjustable, allowing more fine-grained / smoother edges at the cost of processing power. An example of the points generated can be seen on Figure 3.2
3. **Convex hull generation and setting edge size:** Using the concentric point sets, edges are calculated by finding the convex hull of each edge's points with Andrew's monotone chain algorithm. Edge width is adjusted based on the Support value of the frequent patterns. The pseudo code

for Andrew's monotone chain algorithm is presented in Algorithm 1 [43] and the Godot implementation in Appendix A.2

4. **Centrality value calculation and setting node size:** The radius of the nodes are related to the degree centrality calculated as in equation 2.8

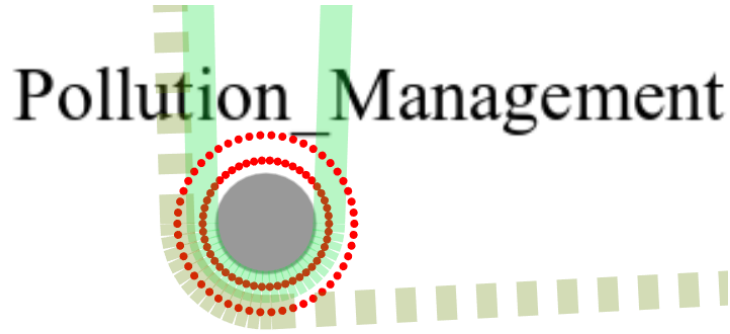


Figure 3.2: Visualization of how the circles and points are generated around a given node, based on the edges

To compare multiple groups the `#GROUP` tag was added as mentioned before. If only 1 group is present (all groups are the same) a classical hypergraph can be seen, but if more groups are defined, each unique group can have a different visualization. Right now 2 groups can be compared, so the first group will have a solid line, and the second group will have a segmented line. If more than 2 groups are added, all subsequent groups will have solid lines. The aim is to provide comparison between up to 5 different hypergraphs in the future.

Algorithm 1 Andrew's Monotone Chain Algorithm

Input: A list P of points in the plane.

Precondition: There must be at least 3 points.

- 1: Sort the points of P by x-coordinate (if there is a tie, sort by y-coordinate).
 - 2: Initialize U and L as empty lists. $\triangleright U$ and L will store the vertices of the upper and lower hulls respectively.
 - 3: **for** $i = 1, 2, \dots, n$ **do**
 - 4: **while** L contains at least two points and the sequence of the last two points of L and $P[i]$ does not make a counter-clockwise turn **do**
 - 5: Remove the last point from L .
 - 6: **end while**
 - 7: Append $P[i]$ to L .
 - 8: **end for**
 - 9: **for** $i = n, n - 1, \dots, 1$ **do**
 - 10: **while** U contains at least two points and the sequence of the last two points of U and $P[i]$ does not make a counter-clockwise turn **do**
 - 11: Remove the last point from U .
 - 12: **end while**
 - 13: Append $P[i]$ to U .
 - 14: **end for**
 - 15: Remove the last point from each list (as it's the same as the first point of the other list).
 - 16: Concatenate L and U to obtain the convex hull of P . $\triangleright$ Points in the result are listed in counter-clockwise order.
-

4. Chapter

Results

The proposed method is demonstrated through an illustrative case study on the Conference of Parties (COP) international sustainability events. Therefore, the following steps were carried out following the workflow indicated in Figure 2.2:

1. Selection of events:

Three of the latest COP events were considered throughout the analysis, namely the 26th Conference of Parties (COP 26) at Glasgow, the 27th Conference of Parties (COP 27) at Sharm El Sheikh, 28th and the Conference of Parties (COP 28) at Dubai. The purpose of these conferences is to bring together countries to negotiate and implement international agreements and policies aimed at combating climate change. The focus of the events is on assessing progress, setting new targets, and coordinating global actions to limit climate change and its impacts.

2. Selection of topics in taxonomic model that proxy the events theme:

The conferences are about climate change, therefore to depict related news, the SDG 13- Climate change-related keywords from the World Bank Topical Taxonomy have been gathered, which includes 49 relevant keywords based on a previous study [12].

3. Selection of time frame for querying news from GDELT:

The news was searched and filtered to include the period during which the event took place, as well as 30 days before and after the event.

4. Defining the time window for frequent itemset mining:

The time window w in the case study was set to one day.

For each event, frequent itemsets are identified within specified time windows. We compare the hypergraphs generated for the whole timeframes.

In Figure 4.1 the comparison between COP26 (Glasgow, 2021) and COP27 (Sharm el-Sheikh, 2022) is shown, while in Figure 4.2 the comparison between COP27 (Sharm el-Sheikh, 2022) and COP28 (Dubai, 2023) is illustrated. Node sizes reflect degree centrality, and shared itemsets between the two datasets have been removed to highlight differences. The individual hypergraphs for each event for the whole timeframe is in

In both figures, topics like "Energy and Extractives," "Climate Change," emerge as central themes with degree centrality 8 for both, indicated by their larger node sizes. This means that these topics involved in 8 hyperedges which are frequent itemset. Over the years, these themes remain critical, suggesting consistent importance across COPs. In Figure 4.2, "Environmental and Natural Resources" appear among the central elements indicating a shift in media central focuses. Furthermore, new connections appear in each subsequent year. For instance, in Figure 4.1, we observe more solid connections related to "Health Emergency Preparedness and Disaster Response" [44] showing prominence in COP26, whereas dashed lines show that these topics may have shifted slightly in COP27, but still remain important.

Hyperedge "Health Emergency Preparedness and Disaster Response, Energy and Extractives, Climate Change" appears with the support value of 4497, which indicated how many times the mentioned itemset appears within news apprearings at the given timeframe. In 2022, hyperedges "Health Emergency Preparedness and Disaster Response, Climate Change" and "Health Emergency Preparedness and Disaster Response, Energy and Extractives" appears with support values 6731 and 7720 accordingly. For 2021 the highest support value is related to "Energy and Extractives, Environment and Natural Resources, Forests" hyperedge with support value 4586. COP27 shows more prominent connections around "Evidence Based Policy" and "Energy Efficiency," suggesting a strong focus on policies to curb climate impacts through specific mitigation strategies [45]. Furthermore, "Evidence Based Policy" have the support value of 5238, indicating a strong focus.

Figure 4.2 indicates the progression from COP27 to COP28. "Environmental and Natural Resources" appears as a more central element relative to

Results

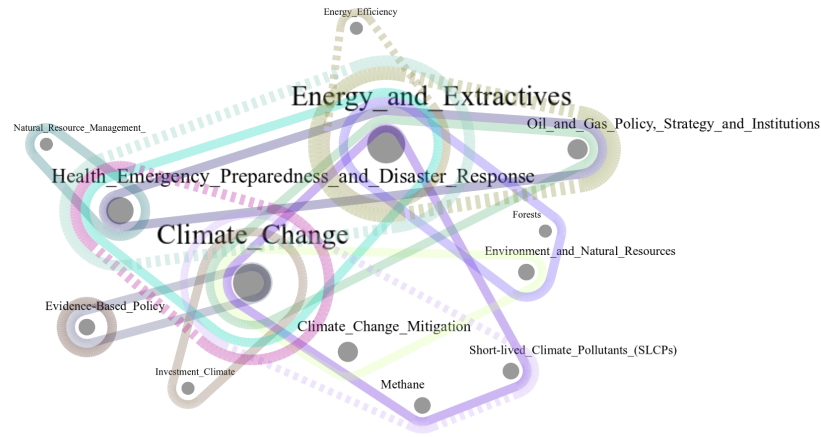


Figure 4.1: Hypergraph comparing frequent itemsets from COP26 (Glasgow, 2021) and COP27 (Sharm el-Sheikh, 2022). Solid lines represent 2021, dashed lines represent 2022. Node sizes reflect degree centrality, and shared itemsets between the two datasets have been removed to highlight differences

previous comparison. For 2022, COP27 "Energy and Extractives, Oil and Gas Policy,Strategy and Institutions" hyperedge has the highest support value fo 14501.

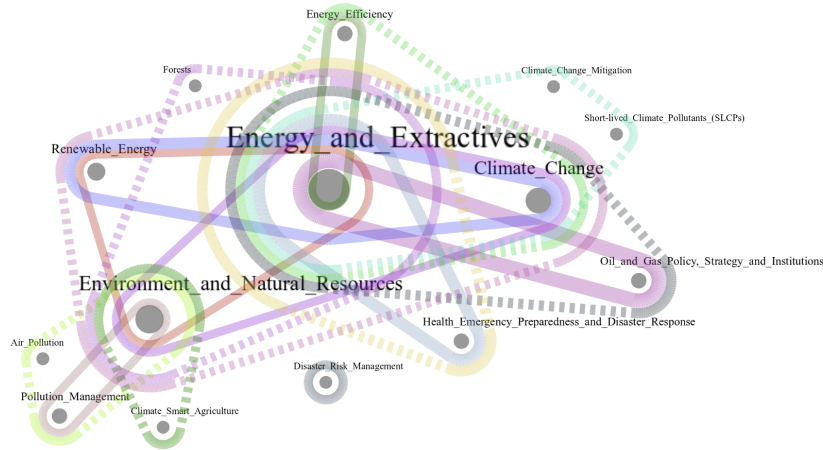


Figure 4.2: Hypergraph comparing frequent itemsets from COP27 (Sharm el-Sheikh, 2022) and COP28 (Dubai, 2023). Solid lines represent 2022, dashed lines represent 2023. Node sizes reflect degree centrality, and shared itemsets have been removed to highlight differences.

While "Energy and Extractives" and "Oil and Gas Policy, Strategy and Institutions" remains a dominant theme, there's a marked increase in connections

around "Renewable Energy" and "Pollution Management" in the dashed lines representing COP28. This itemset suggests an intensified focus in 2023 on renewable energy transition and pollution mitigation, indicating an actionable response to the global call for cleaner energy sources [46]. Additional topics such as "Air Pollution" and "Climate-Smart Agriculture" also gain prominence in COP28, highlighting a shift towards specific environmental and agricultural practices to address climate impacts. These new associations suggest that COP28 discussions may be more granular and sector-specific, with a tangible focus on practical measures. Furthermore, "Disaster Risk Management" with support value 3732 in COP28 implies it became a significant topic, potentially reflecting an urgent need for resilience strategies in light of more frequent climate-related disasters.

In figure 4.3 for each year, the top five most frequent topic itemsets (highest support values) during the conferences are visualized over time. By doing so, we aim to detect whether the popular topics are frequent only during the conference time frame or was it already popular before and after the conference time interval. Therefore, we can gain information about the impact of the conference on media attention. On the x-axis the number of days are shown, while the y-axis indicates the support in percentage as follows:

$$\begin{aligned} \text{supp}_{\text{conf_start}}^{\text{conf_end}}(F_i) &> \text{supp}_{\text{conf_start} - 1 \text{ month}}^{\text{conf_start}}(F_i) \\ &\text{and} \\ \text{supp}_{\text{conf_start}}^{\text{conf_end}}(F_i) &> \text{supp}_{\text{conf_end} + 1 \text{ month}}^{\text{conf_end}}(F_i) \end{aligned} \quad (4.1)$$

, where the `conf_start` is the starting date of the given conference, and `conf_end` is the ending date of the given conference. The start is indicated with green dashed line, while the ending is shown as a red dashed line. The topics are color-coded according to Table 4.1.

The orange trend "Climate Finance, Climate Change" frequent itemset appears each year and tend to peak during the conference intervalls. The blue trend "Climate Change Impacts, Climate Change, Climate Science" frequent itemset happens to be in the top five most popular itemset for COP26 and COP27, but not in COP28.

Overall, we can note that the most popular frequent itemsets are tend to

become important during the conference interval, but less recognised before and after the conference. This indicates that the media attention is more focused regarding these topics than usual.

In 2021 at COP26, the red line "Investment Climate, Climate Change" has a peak in the middle of the conference duration, which aligns to the agenda of COP26, where on day four a greater focus happened on climate finance for mitigation. In news we might experience a shift in its appearing time and follow up. Frequent itemsets characterise the COP26 in the news are as follows: "Health Emergency Preparedness and Disaster Response, Natural Resource Management" (light blue), and "Climate Change Mitigation, Short-lived Climate Pollutants (SLCPs), Methane, Climate Change" (light brown).

In 2022 frequent news itemsets are characterised around: "Energy and Extractives, Renewable Energy, Environment and Natural Resources" (light green), and "Energy and Extractives, Environment and Natural Resources, Climate Change" (brown). Each news tend to become important during the conference interval, indicating a media focus on related topics.

Finally, in 2023 the most popular frequent news topics include: "Oil and Gas Policy, Strategy and Institutions, Energy and Extractives, Climate Change" (light red), "Energy and Extractives, Renewable Energy, Energy Efficiency, Climate Change" (purple), and "Climate Smart Agriculture, Climate Change" (light purple). "Environment and Natural Resources, Natural Resource Management, Climate Change" (light orange) only show a slight increase at the second half of the conference.

Results

Itemset	Color
Climate Change Impacts, Climate Change, Climate Science	
Health Emergency Preparedness and Disaster Response, Natural Resource Management	
Climate Finance, Climate Change	
Environment and Natural Resources, Natural Resource Management, Climate Change	
Natural Resources Law	
Energy and Extractives, Renewable Energy, Environment and Natural Resources	
Investment Climate, Climate Change	
Oil and Gas Policy, Strategy and Institutions, Energy and Extractives, Climate Change	
Energy and Extractives, Renewable Energy, Energy Efficiency, Climate Change	
Climate Smart Agriculture, Climate Change	
Energy and Extractives, Environment and Natural Resources, Climate Change	
Climate Change Mitigation, Short-lived Climate Pollutants (SLCPs), Methane, Climate Change	

Table 4.1: Itemsets with associated colors for trends.

Results

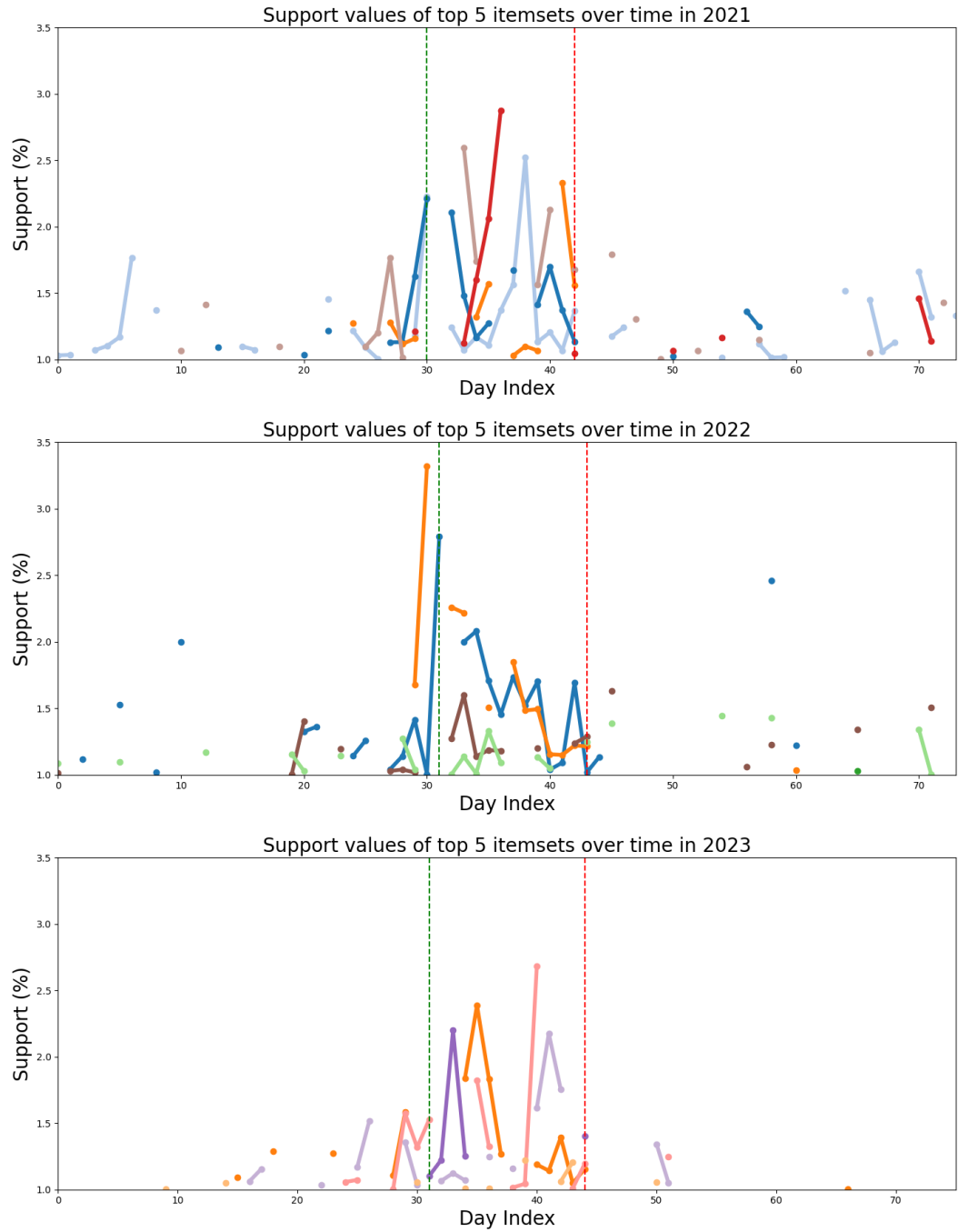


Figure 4.3: Change of the support values of the frequent itemsets which became hot topics during the COP in 2021, 2022 and 2023

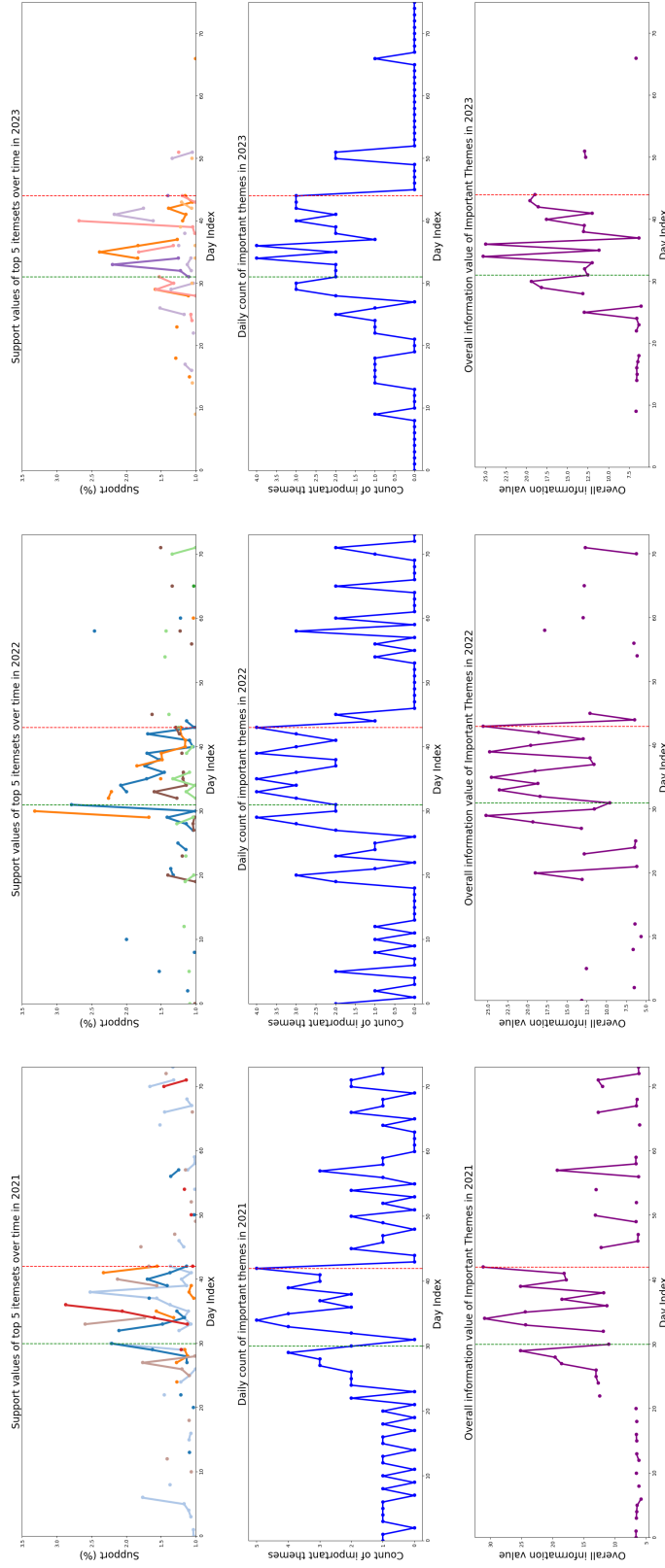


Figure 4.4: Comparison of the hot topics (important frequent itemsets) support values, itemset counts, and information values of COP conferences held in 2021, 2022, and 2023

Finally, Figure 4.4 provides a comparative analysis of the top five frequent itemsets identified in Figure 4.3 for each year, showcasing the main news topics related to COP26, COP27, and COP28 across 2021, 2022, and 2023. These top frequent itemsets are considered as hot topics gaining popularity during the conferences. Each column represents a different year, while the three rows emphasize various aspects of media coverage trends: 1) support of hot topics in % (top row), 2) frequent itemsets (hot topics) count (middle row), 3) sum of the hot topics' information value (bottom row). The dashed lines in each subfigure represents the start (green) and finish (red) date of the COP conferences.

The synchronized growth of support, frequent itemset count, and information value across COP events suggests that as each conference approaches, media attention intensifies both in volume and in the importance of frequent news topics covered. This itemset implies that the media not only increases coverage on COP-related themes but also prioritizes more impactful narratives, reflecting heightened public and political interest. Lower information values signal more focused media attention on select topics, while higher values indicate a broader range of themes covered.

The three rows together show a cycle of media focus: a buildup phase where a few key topics gain traction, a peak during the conference with a broader range of high-value themes, and a post-conference decline, indicating a return to normal coverage levels. This approach underscores how media coverage expands and contracts around COP events, shaping media focus, and emphasizing influential hot topics during the peak periods.

5. Chapter

Conclusion

This study demonstrated a method for measuring the impact of specific events on media coverage and provided a robust tool for analyzing frequent itemsets from subsequent timeframes. A novel hypergraph visualization tool was developed to enhance the accessibility of the results. The presented approach can accurately identify key topics of discussion during international events. The methodology was applied to data from the Global Database of Events, Language, and Tone (GDELT), during the COP 2021, 2022 and 2023. After collecting and processing the data, frequent pattern mining was used to identify the frequent itemsets for each timeframe, which is used to construct hypergraphs to showcase the frequent topics of discussion for that timeframe. The methodology allows for the comparison of frequent itemsets across consecutive events, highlighting the changes in topics, showing the shifts in media focus. By focusing on the differences of frequent patterns from different timeframes, temporal hypergraphs were constructed to visualize the changes in topics of discussion from 2021 to 2022 and from 2022 to 2023. This approach provides a clear and robust way of following how the topics changed over the time, making it possible to track the evolving media landscape surrounding critical issues.

6. Chapter

Future plans

The next steps involve expanding the visualization tool with enhanced customization and performance optimizations, enabling better handling of larger datasets and more interactive analysis. I also plan to publish a scientific paper to share the methodology and findings. Additionally, I plan to utilize the acquired knowledge in my Master Thesis, where I plan to integrate Deep Learning techniques into the presented methodology to analyze data more effectively and incorporate external variables, such as GDP or happiness indexes, to provide deeper insights. I plan on expanding the visualization tool with a simulation space, capable of communicating with external analysis, specifically showcasing changes in time in a more user-friendly way. Finally, efforts will be made to develop a model that links events, visualizing the chain of events leading up to a specific outcome. Based on the news-related analysis projects and results developed within the University of Pannonia under the Multidisciplinary National Laboratory for Climate Change (RRF-2.3.1-21-2022-00014), an interactive webpage is being developed that provides a user-friendly way to interact with the data provided by research. The developed methodology will be integrated into the website to expand its functionality.

Bibliography

- [1] United Nations. Global issues. <https://www.un.org/en/global-issues> **last accessed:** [2024-06-06].
- [2] Kalterina Shulla and Walter Leal-Filho. Achieving the un agenda 2030: Overall actions for the successful implementation of the sustainable development goals before and after the 2030 deadline, 2023.
- [3] Tomas Pernecky and Michael Lück. *Events, society and sustainability*. Taylor & Francis Group. Routledge, 2012.
- [4] Ralf Barkemeyer, Frank Figge, and Diane Holt. Sustainability-related media coverage and socioeconomic development: A regional and north-south perspective. *Environment and Planning C: Government and Policy*, 31(4):716–740, 2013.
- [5] GDELT. World bank group topical taxonomy now in gkg. <https://blog.gdeltproject.org/world-bank-group-topical-taxonomy-now-in-gkg/> **last accessed:** [2024-06-06].
- [6] GDELT. Crisislex taxonomies now available in gkg. <https://blog.gdeltproject.org/crisislex-taxonomies-now-available-in-gkg/> **last accessed:** [2024-06-06].
- [7] Kathleen Buckingham, John Brandt, Will Anderson, Luiz Fernando do Amaral, and Ruchika Singh. The untapped potential of mining news media events for understanding environmental change. *Current Opinion in Environmental Sustainability*, 45:92–99, 2020.

BIBLIOGRAPHY

- [8] Jing Chen, Germán G Creamer, Yue Ning, and Tal Ben-Zvi. Healthcare sustainability: hospitalization rate forecasting with transfer learning and location-aware news analysis. *Sustainability*, 15(22):15840, 2023.
- [9] Markus Bertl. News analysis for the detection of cyber security issues in digital healthcare: A text mining approach to uncover actors, attack methods and technologies for cyber defense. *Young Information Scientist*, 4:1–15, 2019.
- [10] Alberto Barroso del Toro. Understanding investor reaction to sustainability news in energy companies: fossil vs renewable vs nuclear, tone effects, and covid-19 impacts through a multi-study approach using event study methodology. 2023.
- [11] Vasiliki Voukelatou, Luca Pappalardo, Ioanna Miliou, Lorenzo Gabrielli, and Fosca Giannotti. Estimating countries’ peace index through the lens of the world news as monitored by gdelt. In *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*, pages 216–225. IEEE, 2020.
- [12] Timea Czvetko, Gergely Honti, Viktor Sebestyen, and Janos Abonyi. The intertwining of world news with sustainable development goals: An effective monitoring tool. *Heliyon*, 7(2), 2021.
- [13] Yunxing Yao, Yinbao Zhang, Jianzhong Liu, Yanpei Li, and Xiaopei Li. Analysis of spatiotemporal characteristics and influencing factors for the aid events of covid-19 based on gdelt. *Sustainability*, 14(19):12522, 2022.
- [14] Rafael Prieto Curiel, Carmen Cabrera Arnau, Mara Torres Pinedo, Humberto González Ramírez, and Steven R Bishop. Temporal and spatial analysis of the media spotlight. *Computers, Environment and Urban Systems*, 75:254–263, 2019.
- [15] Donna J Peuquet, Anthony C Robinson, Samuel Stehle, Franklin A Hardisty, and Wei Luo. A method for discovery and analysis of temporal patterns in complex event data. *International Journal of Geographical Information Science*, 29(9):1588–1611, 2015.

BIBLIOGRAPHY

- [16] Guoshuai Zhang, Jiaji Wu, Mingzhou Tan, and Hong Han. Predicting social events with multimodal fusion of spatial and temporal dynamic graph representations. *Big Data*, 10(5):440–452, 2022.
- [17] C Piselli, A Fronzetti Colladon, Ludovica Segneri, and Anna Laura Pisello. Evaluating and improving social awareness of energy communities through semantic network analysis of online news. *Renewable and Sustainable Energy Reviews*, 167:112792, 2022.
- [18] Mehedi Tajrian, Azizur Rahman, Muhammad Ashad Kabir, and Md Rafiqul Islam. A review of methodologies for fake news analysis. *IEEE Access*, 2023.
- [19] Hyun Park, Jinie Pak, and Yanggon Kim. Analysis of news article various countries on a specific event using semantic network analysis. In *International Conference on Information Technology-New Generations*, pages 229–235. Springer, 2023.
- [20] Anoushka Harit, Zhongtian Sun, Jongmin Yu, and Noura Al Moubayed. Breaking down financial news impact: A novel ai approach with geometric hypergraphs. *arXiv preprint arXiv:2409.00438*, 2024.
- [21] Yunshan Ma, Chenchen Ye, Zijian Wu, Xiang Wang, Yixin Cao, and Tat-Seng Chua. Context-aware event forecasting via graph disentanglement. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1643–1652, 2023.
- [22] A Garaeva, F Makhmutova, I Anikin, and Kai-Uwe Sattler. A framework for co-location patterns mining in big spatial data. In *2017 XX IEEE International Conference on Soft Computing and Measurements (SCM)*, pages 477–480. IEEE, 2017.
- [23] Youcef Djenouri, Asma Belhadi, Gautam Srivastava, and Jerry Chun-Wei Lin. Advanced pattern-mining system for fake news analysis. *IEEE Transactions on Computational Social Systems*, 10(6):2949–2958, 2023.
- [24] Mauricio Daniel Quezada Veas. Knowledge discovery from news events on twitter. 2020.

BIBLIOGRAPHY

- [25] GDELT. Querying, analyzing and downloading. <https://www.gdeltproject.org/data.html> last accessed: [2023-10-31].
- [26] Rumeng Chen, Feng Hu, Feng Wang, and Libing Bai. Hypergraph-clustering method based on an improved apriori algorithm. *Applied Sciences*, 13(19):10577, 2023.
- [27] Alain Bretto. Hypergraph theory. *An introduction. Mathematical Engineering. Cham: Springer*, 1, 2013.
- [28] Joshua Pickard, Can Chen, Rahmy Salman, Cooper Stansbury, Sion Kim, Amit Surana, Anthony Bloch, and Indika Rajapakse. Hat: Hypergraph analysis toolbox. *PLOS Computational Biology*, 19(6):e1011190, 2023.
- [29] H. Ackermann P. Purkait, T.J. Chin and D. Suter. Clustering with hypergraphs: The case for large hyperedges. In *European Conf. on Computer Vision (ECCV)*, 2014.
- [30] Marcos Bolanos. Hypergraph edge/vertex matrix. 2023.
- [31] Brenda Praggastis, D Arendt, C Joslyn, E Purvine, S Aksoy, and K Monson. Hypernetx, 2019.
- [32] Vaibhav Karve. Multihypergraph.
- [33] Youjia Zhou, Archit Rathore, Emilie Purvine, and Bei Wang. Topological simplifications of hypergraphs. *IEEE Transactions on Visualization and Computer Graphics*, 2022.
- [34] Quintino Francesco Lotito, Martina Contisciani, Caterina De Bacco, Leonardo Di Gaetano, Luca Gallo, Alberto Montresor, Federico Musciotto, Nicolò Ruggeri, and Federico Battiston. Hypergraphx: a library for higher-order network analysis. *Journal of Complex Networks*, 11(3):cnad019, 2023.
- [35] Brendan Avent, Anna Ritz, T Murali, J Cadena, and Y Keneshloo. Halp: Hypergraph algorithms package. 2014.
- [36] David J. Marchette. Hyperg: Hypergraphs in r. 2021.

BIBLIOGRAPHY

- [37] George Karypis and Vipin Kumar. A hypergraph partitioning package. *Army HPC Research Center, University of Minnesota, Department of Computer Science & Engineering*, 1998.
- [38] Kuldeep Kurte, Neena Imam, SM Shamimul Hasan, and Ramakrishnan Kannan. Phoenix: A scalable streaming hypergraph analysis framework. In *Advances in Data Science and Information Engineering: Proceedings from ICDATA 2020 and IKE 2020*, pages 3–25. Springer, 2021.
- [39] Alexis LG. Hypergraphlib.
- [40] Léo PM Diaz and Michael PH Stumpf. Hypergraphs. jl: representing higher-order relationships in julia. *Bioinformatics*, 38(14):3660–3661, 2022.
- [41] Louis Jenkins, Tanveer Bhuiyan, Sarah Harun, Christopher Lightsey, David Mentgen, Sinan Aksoy, Timothy Stavenger, Marcin Zalewski, Hugh Medal, and Cliff Joslyn. Chapel hypergraph library (chgl). In *2018 IEEE high performance extreme computing conference (HPEC)*, pages 1–6. IEEE, 2018.
- [42] Godot Engine. <https://github.com/godotengine/godot/releases/tag/4.3-stable> **last accessed:** [2023-10-10].
- [43] WikiBooks. Algorithm implementation/geometry/convex hull/monotone chain. https://en.wikibooks.org/wiki/Algorithm_Implementation/Geometry/Convex_hull/Monotone_chain **last accessed:**[2024-11-07].
- [44] Martina Valente, Monica Trentin, Luca Ragazzoni, and Francesco Barone-Adesi. Aligning disaster risk reduction and climate change adaptation in the post-cop26 era. *The Lancet Planetary Health*, 6(2):e76–e77, 2022.
- [45] Samuel Asumadu Sarkodie, Maruf Yakubu Ahmed, and Phebe Asantewaa Owusu. Advancing cop26 climate goals: Leveraging energy innovation, governance readiness, and socio-economic factors for enhanced climate resilience and sustainability. *Journal of Cleaner Production*, 431:139757, 2023.

- [46] Pranay Arora. Cop28: ambitions, realities, and future. *Environmental Sustainability*, 7(1):107–113, 2024.

A. Chapter

Appendix

A.1 Example data accepted by HyperGodot

```
1 Environment_and_Natural_Resources Pollution_Management #SUP:
  2662 #GROUP 2022
2 Energy_and_Extractives Energy_Efficiency #SUP: 3163 #GROUP 2022
3 Energy_and_Extractives
  Health_Emergency_Preparedness_and_Disaster_Response #SUP:
  7720 #GROUP 2022
4 Environment_and_Natural_Resources Pollution_Management
  Air_Pollution #SUP: 3312 #GROUP 2023
5 Environment_and_Natural_Resources Energy_and_Extractives
  Climate_Change Renewable_Energy #SUP: 3301 #GROUP 2023
```

A.2 Andrew's monotone chain GDScript implementation

Input: The points around the nodes that are inside the given edge.

Output: The points in order we need to draw lines to. (We draw a line segment from point x0 to point x1, from x1 to x2 etc..)

```
1 func andrew_monotone_chain(points):
2   if points.size() < 3:
3     return points #3 points always define a convex hull
4   # Sort points based on their x coordinate first, and y
    coordinate second
5   points.sort_custom(func(a, b):
6     return a.x < b.x if a.x != b.x else a.y < b.y
7   )
8   # Lower hull
9   var lower = []
```

```

10 for p in points:
11     while lower.size() >= 2 and cross(lower[lower.size() - 2],
12         lower[lower.size() - 1], p) <= 0:
13         lower.pop_back()
14     lower.append(p)
15 # Upper hull
16 var upper = []
17 for i in range(points.size() - 1, -1, -1):
18     var p = points[i]
19     while upper.size() >= 2 and cross(upper[upper.size() - 2],
20         upper[upper.size() - 1], p) <= 0:
21         upper.pop_back()
22     upper.append(p)
23 # Concatenate the hulls, but the first and last points are the
24     same for the two hulls, so we pop them.
25 lower.pop_back()
26 upper.pop_back()
27 return lower + upper
28
29 func cross(o, a, b):
30     return (a.x - o.x) * (b.y - o.y) - (a.y - o.y) * (b.x - o.x)

```