



CO-WORD GRAPHS FOR DESIGN AND MANUFACTURE KNOWLEDGE MAPPING

J. Gopsill^{1,✉}, M. Humphrey², D. Thompson² and E. Garcia²

¹ University of Bath, United Kingdom, ² National Composites Centre, United Kingdom

✉ J.A.Gopsill@bath.ac.uk

Abstract

Design & Manufacture Knowledge Mapping is a critical activity in medium-to-large organisations supporting many organisational activities. However, techniques for effective mapping of knowledge often employ interviews, consultations and appraisals. Although invaluable in providing expert insight, the application of such methods is inherently intrusive and resource intensive. This paper presents word co-occurrence graphs as a means to automatically generate knowledge maps from technical documents and validates against expert generated knowledge maps.

Keywords: knowledge management, technology development, design informatics, graph theory

1. Introduction

Design & Manufacture (D&M) Knowledge Mapping is a critical activity in medium-to-large organisations. The results support many organisational activities, such as organisational structuring, resource allocation, risk assessment and organisational strategy. Current techniques for effective mapping of knowledge employ interviews, consultations and appraisals (Jafari et al., 2009). Although invaluable in providing expert insight into an organisation, the application of such methods is inherently intrusive and resource intensive as they require significant time from experts to gather the necessary data and subsequent post-processing to synthesise the knowledge map. The level of intrusion required also scales with the size of the organisation and leads to knowledge mapping activities taking months or even years to complete. For example, the generation of a knowledge map for a timber construction company was achieved through an iterative interview and survey involving employees working on 60 timber structure projects over a six-month period (Björnfort and Stehn, 2007).

The mapping of knowledge becomes even more challenging when one considers Research & Development (R&D) organisations where new knowledge is being generated on a daily basis. This can quickly lead to out-dated maps due to emergent terms and topics. In addition, global organisations have R&D facilities distributed across the world that have their own domain expertise, which poses challenges in the awareness and sharing of knowledge across transdisciplinary organisations and projects.

In these organisations, technical reporting remains the main form of knowledge capture and the advent of digital document management systems has provided an accessible store that could be automatically analysed by data mining algorithms to extract insights into the knowledge of the organisation. However, the application of ‘off-the-shelf’ industry search technologies has been unable to effectively index and categorise these reports. Jones et al. (2015) investigation into the search & retrieval behaviour of employees in Airbus revealed that employees performed 1.1 M searches with many

having to perform extended repeat searches to find the information they required. This has been attributed to the lexical and logical semantic of technical reports, which is unlike websites, social media and newspapers that the search technologies have been trained on. Thus, new approaches that are able to learn the semantics and identify domain-specific keywords is required, while also providing user-in-the-loop validation that gives employees confidence in the results being generated.

This paper contributes to the field through an adaptation of co-word analysis that enables the process to be applied to full-text documents to generate a graph of connected terminology. It is posited, (and investigated in this paper,) that the resulting term graph properties can be used to determine whether the term should be considered a knowledge element through supervised learning where experts have labelled a sub-set of terms in the graph. Achieving this would provide an automated method of determining emergent knowledge elements within technical document repositories.

Having identified the knowledge elements, the graph can be filtered to provide a sub-graph representing the knowledge in an organisation. This can be partitioned to identify topics that are then mapped to designers based on authorship and document access. The effectiveness of a community partitioning algorithm is examined and compared to partitioning by a group of discipline experts.

The paper continues by detailing the related work in the fields of knowledge mapping and co-word analysis (Section 2). This is followed by the approach that adapts co-word analysis to the generation of D&M knowledge maps that feature emergent knowledge elements (Section 3). To validate the approach, a scoping study to validate the approach and determine whether graph properties can be used to identify knowledge elements and subsequent partitioning into topics to the same degree of confidence as experts has been performed (Section 4). The findings are then discussed along with an outline of future work (Section 5). The paper concludes by highlighting the key findings from the scoping study (Section 6).

2. Related work

This section provides a summary of the state-of-the-art in the fields of knowledge mapping and application of co-word analysis.

2.1. Knowledge mapping

Knowledge mapping is a key task for organisations and a number of approaches have been developed and implemented in industry. [Jafari et al. \(2009\)](#) review of knowledge mapping identified 5 techniques used in practice: Yellow Paging, Information Flow Analysis, Social Network Analysis, Process Knowledge Mapping and Functional Knowledge Mapping. The comparison of these techniques (Table 1) reveals that the majority of techniques are laborious and require considerable resource from the organisation to complete.

In design research, researchers have been exploring how one can overcome the laborious nature of current industry standard techniques. For example, [Dori and Shpitalni \(2005\)](#) has sought to define an ontology of knowledge through extraction of explicit links in Product Data Management (PDM) systems. While [Sarica et al. \(2019\)](#) and [Sarica et al. \(2020\)](#) have investigated patent databases as a consistent and reliable source to synthesise and organise knowledge on a subject matter. They evaluated co-word analysis, part-of-speech tagging and a graph-based unsupervised extraction method. [Schmidt et al. \(2015\)](#) mapping of knowledge in student design projects utilised pre-defined knowledge elements that were then related to elements of the project and design process.

Having generated a map, considerable research has been performed to exploit them to generate insights for industry. Example analyses include, Multi-Domain Matrices, Structural Criteria and Domain Mapping Matrices, being applied to support decision-making, project planning, strategy development and risk management ([Schmidt et al., 2013b](#); [Schmidt et al., 2013a](#); [von Saucken et al., 2014](#)).

This brief summary of knowledge mapping highlights that industry is still using manual and laborious processes to generate knowledge maps. However, recent advancements in design research has highlighted the potential of the digital footprint, such as PDM systems and document repositories, as a data source for the generation of knowledge maps through numerical techniques. Yet, it is the validation and confidence in the results that appears to be the continuing barrier in pulling this

research through into industry. In addition, the challenge of maintaining up-to-date knowledge maps where the document set is continually evolving, such as in R&D, also remains.

Table 1. Comparison of knowledge mapping techniques (Jafari et al., 2009)

	Yellow Page	Information Flow Analysis	Social Network Analysis	Process Knowledge Mapping	Functional Knowledge Mapping
Used tools for data gathering.	Question and answer systems, Skill dictionary and reports.	Interviews, skill inventories, and extensive surveys. Information flow diagrams (IFD).	Questionnaire, Sociogram and Graph Theory.	Brainstorming or conduct interviews with the process owners.	Survey and interview.
Used tools for knowledge map evaluation.	Skills dictionary.	Questionnaire, interviews and sign-out sheets.	InFlow, Krackplot and NetMiner.	–	Observations, Interviews, Internal Reports.
Objectives.	Create transparency as to the location of knowledge in the organisation by registering individual competencies in a database or similar manner.	Determining who is accessing what information sources and how often.	Discover interaction patterns between members.	Define the knowledge needed, decision milestones, the knowledge available to support a business process. Routes for access and retrieval of knowledge. Gaps between required skills and current skills.	Locate knowledge sensitive areas. Identifies and characterises areas of process related critical knowledge spots.
Knowledge mapping approach	Project-based	Relationship-based	Relationship-based	Process-based	Process-based
Create static or dynamic map	Static	Static	Dynamic	Dynamic	Dynamic
Support tacit or explicit	Explicit	Tacit	Tacit	Explicit, Tacit	Explicit, Tacit

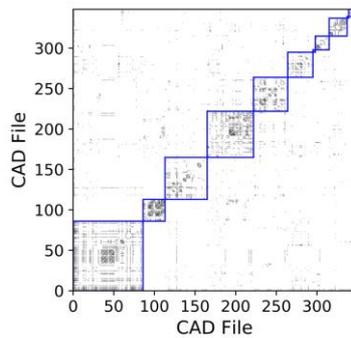
2.2. Co-word analysis

Co-word analysis evaluates document repositories through the co-occurrence of terms. By analysing the co-occurrence of terms, a graph of connected terms (a.k.a vertices in graph theory) is generated, which enables the application of graph theory algorithms to uncover emergent structures and patterns between terms. For example, centrality measures are often used to identify the most important and influential terms, while partitioning algorithms, such as Louvain community partitioning (Blondel et al., 2008; Hagberg et al., 2008), seeks to identify groups of highly connected terms within a graph, where Modularity (Q) is used as the objective function (Newman 2006; Newman 2004). In the context of co-word analysis, the partitioning of terms is used to identify themes/topics.

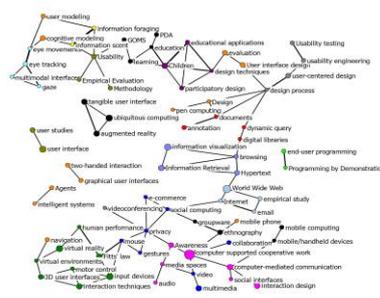
Co-word analysis has primarily been applied in the identification of research themes within scientific communities using keywords, abstracts and summaries of documents to identify the terms of interest (Ding et al., 2001; Liu et al., 2014). It has also been used in document sets where the text has been broken down to the sentence level (Le and Mikolov 2014). While Gopsill et al. (2015) has applied co-word analysis to small text, such as engineering e-mail, enabling the evolution of topics to be monitored and associations with requirements scope creep to be developed. Small texts have also been of interest to He et al. (2019) who has applied co-word analysis to synthesise concept designs from crowd-sourced ideation exercises.

It is not only the quantitative metrics afforded by this analytical technique but also the ability to aggregate and visualise a large corpus of information into more manageable forms for human interpretation and decision making (Figure 1). A number of visualisation techniques have been employed and include:

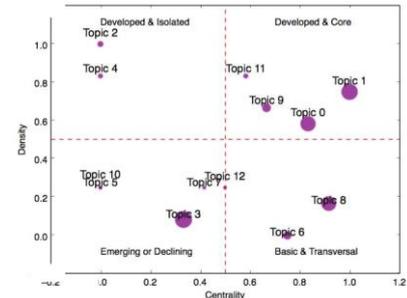
- Re-arranged matrices in relation to the partitioning of terms;
- Force-based network diagrams to reveal the connected nature of the terms; and,
- Strategic diagrams that show the movement of topics over time.



(a) Matrix view
(From: Gopsill et al., 2016)



(b) Network diagram
(From: Liu et al., 2014)



(c) Strategic diagram
(From: Jones et al., 2015)

Figure 1. Graph visualisations

From this brief review of the state-of-the-art, the challenge identified is in generating accurate and detailed knowledge maps through non-information foraging means. A challenge that could be overcome with an approach that moves co-word analysis away from pre-defined keywords and to whole-body text.

3. Co-word knowledge maps

The related work highlights the potential for co-word analysis to complement existing knowledge mapping approaches and one that would be more suitable for large organisations with vast repositories of technical reports. This section proposes how co-word analysis, which has typically been used on keywords and short texts, could be adapted and used to identify keywords from full-text and subsequent partitioning of these terms to identify topics of knowledge for an organisation.

3.1. Graph generation

The proposed process for generating co-word graphs from full-text technical engineering reports (in portable document format, PDF) follows a 4-step process:

1. Document text extraction;
2. N-gram identification;
3. N-gram co-occurrence; and,
4. Co-occurrence normalisation.

These are now discussed.

Step 1 requires parsing of the PDFs and extraction of the text in UTF-8 format (Step 1). Most technical documents are digitally curated, however if scanned, Optical Character Recognition (OCR) is required. Regular expressions are then used to extract the year in which the report was created, authorship, and generate a list of noun-noun bi-grams contained within the main body of document (Step 2). Noun-noun bi-grams have been selected as the objective is to identify technical terms, which are often formed of multiple nouns combined together (e.g. carbon fibre). It is also hypothesised that the noun-noun bi-grams are the most stable feature of the lexicon and thus, suitable for knowledge mapping. The noun-noun bi-grams are then stemmed and checked against a noun stop-word list. Any bi-grams featuring a stop-word are removed and the list of bi-grams is reduced to form a unique set. The noun-noun bi-grams form the graph vertices whose frequency (f) is determined by their appearance in the document set and document frequency (f_d) determined by number of documents the term appears in.

To form the co-word edges (Step 3), indexing of the noun-noun bi-grams is performed for each document and an ordered list of bi-grams with their position indices produced. One can then loop through the list and based on a look-ahead distance, d , determine the bi-grams that co-occur. d or the re-appearance of the term is used as the maximum distance to prevent double counting. With the co-occurrences identified, the process either creates an edge if one does not exist or increments the weight of the edge if it does exist. Figure 2 outlines the approach through a logic-flow diagram.

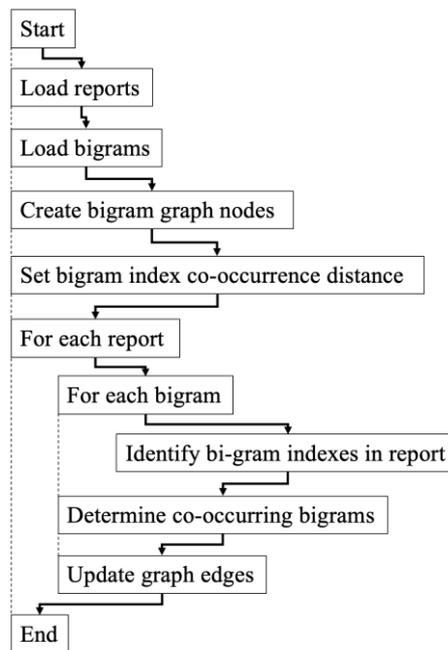


Figure 2. Logic-flow for determining the co-occurrence of bi-grams

Applying a look-ahead distance accounts for the likelihood of multiple topics being discussed across the sections and paragraphs of text within the technical document. d was set to 140 characters to follow examples in social media, such as twitter, where the average sentence length is 100 characters with the extra characters enabling the use of #tags. In the case for technical documents, the additional length in the threshold accounts for topics discussed across multiple sentences and paragraphs.

To account for variations in bi-gram frequency, normalisation is performed:

$$n_{i,j} = \frac{w_{i,j}}{(f_i + f_j - 1)} \quad (1)$$

Where $n_{i,j}$ is the normalised edge weighting and is determined by the number of times i and j have co-occurred ($w_{i,j}$) divided by the number of co-occurrences that could have occurred ($f_i + f_j - 1$). With the co-word graph generated, one can begin to use graph properties to determine keywords and prune on the keywords to form a subgraph and hence, knowledge map.

3.2. Graph properties for keyword identification

The affordance of forming a co-word graph is in the provision of a number of properties that can support us in determining whether a noun-noun bi-gram accurately reflects an element of knowledge. There is the frequency, f , and document frequency, f_d . In addition, one can look at the degree and weighted degree of the bi-gram to inform us on how connected the bi-gram is with the all other bi-grams in the graph. Centrality measures, such as degree, eigenvector and betweenness, provide an alternative perspective on how central a bi-gram is to the graph. With these properties afforded to us, there is a question as to whether they could be used to determine noun-noun bi-grams that reflect elements of knowledge within an organisation.

3.3. Graph partitioning for topic identification

To identify topics through term co-occurrence, the Louvain partitioning algorithm has been applied (Blondel et al., 2008). The algorithm works by allocating terms to partitions with the objective of returning a partition set that produces the highest modularity score for the graph. Modularity (Q) is an assessment of the quality of the matrix partition and is defined as (Newman, 2004):

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (2)$$

Where $m = \frac{1}{2} \sum_{ij} A_{ij}$ and is the number of co-occurrences within the graph. δ is the Kronecker delta function and is 1 if a co-occurrence exists between two terms and 0 otherwise. $k_i k_j / 2m$ is the probability that a co-occurrence may exist between two terms, where k_i is the number of terms that have co-occurrences with term i , and k_j is the number of terms that have co-occurrences with term j . And, A_{ij} is the weighted co-occurrence between two terms in the graph.

To start, the algorithm assigns each term to its own partition. The algorithm sequentially moves one file to a different partition and calculates the change in Q (Equation 3).

$$\Delta Q = \left[\frac{P_{in} + k_{i,in}}{2m} - \left(\frac{P_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{P_{in}}{2m} - \left(\frac{P_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (3)$$

Where P_{in} is sum of all the normalised co-occurrence weights within the partition that term i is being included in, P_{tot} is the sum of all the normalised co-occurrence weights to term within the partition that i is being included in, k_i is the co-occurrence degree of i , $k_{i,in}$ is the sum of the normalised co-occurrence weights i and other terms within the partition that i is merging with.

From this, ΔQ_{max} can be identified. The associated terms are then assigned to the same partition and the algorithm repeats the previous step of identifying the next term movement that will result in a further increase in modularity. If no further movement of terms achieves an increase in modularity, the algorithm terminates resulting in the partition set that gives the highest modularity score. This iterative process results in a partition set that are highly connected internally and weakly connected to one another.

Hence, it is a form of hierarchical clustering and the algorithm iterates until the modularity can no longer be increased by further aggregation of the terms. This paper uses the community API implementation of the Louvain community partitioning algorithm within the NetworkX python package¹.

4. Study: Achieving confidence through expert supervised learning

To test the approach, a scoping study was performed on a subset of the National Composites Centre's (NCC) technical report corpus. The National Composites Centre (NCC) is a world-leading R&D centre for UK composites. Established in 2009 as a result of the UK Composite strategy, it is now part of the UK government's CATAPULT programme of world-leading centres designed to accelerate the UK's High-Value Manufacturing capability and drive future economic growth. The NCC currently provides R&D support for over 40 companies.

While the full technical report repository consists of over 5,000 reports, the subset of reports used in the scoping study consisted of public-funded research projects over a four-year period. Details of the subset used is provided in Table 2.

Table 2. Dataset Statistics

Year	# Documents	Σ Words	Mean # Words	Max. # Words	Min. # Words
2012	5	31 941	6388	9396	3581
2013	6	56 201	9367	17 752	2810
2014	7	83 776	11 968	20 375	999
2015	8	68 107	8513	18 999	4022
Combined	26	240 025	9232	20 375	999

Following the process outlined in Section 3 the co-word bi-gram graph generated featured 23,869 vertices and 130,000 edges. This graph was further pruned based on:

- low document frequency vertices ($f_d < 3$);
- high document frequency vertices ($f_d > 23$);
- low normalised edge weights ($e_w < 0.1$); and,
- low co-occurrence frequency ($e_c < 3$);

resulting in a graph featuring 834 vertices and 1,230 edges.

¹ <https://bitbucket.org/taynaud/python-louvain/src/default/>

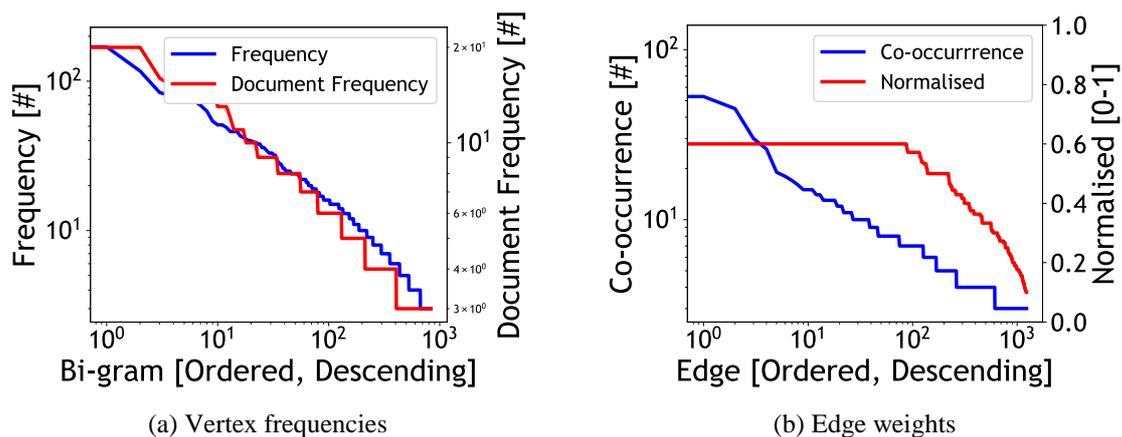


Figure 3. Bi-gram co-word graph properties

4.1. Graph properties of keywords

With the bi-gram co-word graph pruned, the bi-grams were extracted and placed into Excel. The spreadsheet was given to a group of 5 experts who collectively determined which bi-grams were keywords representing elements of knowledge within the NCC. This resulted in a list of 104 keywords from the 354 vertices provided ($\sum x_k = 104, \sum x_{nk} = 250$).

To determine whether the graph properties could be used to differentiate between keywords and non-keywords, T-tests were performed on five properties of the graph: bi-gram frequency, bi-gram document frequency, degree, weighted degree and eigenvector centrality (λ). The results, shown in Table 3, highlight a significance difference in distributions is observed for all but bi-gram frequency and document frequency. This suggests that confidence bounds based on the distributions could be placed on the graph properties to distinguish between keyword and non-keyword bi-grams. Reviewing \bar{x}_k & \bar{x}_{nk} reveals that keywords appear less across the documents and have a lower degree, weighted degree and centrality.

Table 3. T-test of keyword/non-keyword properties

Property	\bar{x}_k	\bar{x}_{nk}	p
Frequency	12.33	9.28	0.06
Document Frequency	4.30	4.23	0.82
Degree	1.2	3.01	< 0.001
Weighted Degree	0.26	0.97	< 0.001
λ Centrality	0.00002	0.0091	< 0.001

4.2. Topic identification

With the keywords identified, the graph was pruned so that only these terms remained with the additional criteria that they needed to be connected in the graph (i.e. no degree of 0). This left 87 terms in the graph. The graph was subsequently partitioned using the Louvain partitioning algorithm (Section 3.3), which gave 17 partitions (10 of which are components) and $Q = 0.836$ (Figure 4a). This score reveals that significant structure is present within the graph as it is generally accepted that $Q > 0.3$ indicates a structure beyond that of random association (Newman, 2004).

To investigate the potential of co-word partitioning as a means to generate knowledge maps, the group of experts were tasked to partition the keywords into areas of knowledge (Figure 4b). The experts partitioned the keywords into 9 partitions.

To compare partitions, the edit distance was applied between the two results as well as comparison with random partitioning of the keywords (Deibel et al., 2005). Taking the Louvain partitioned set, L , and Expert partitioned set, E , the edit distance reflects the minimum number of

moves that are required to convert L to E . This can be solved by determining the maximum number of terms not moved and can be found through maximum weight matching on a bipartite graph. A bipartite graph can be formed on partitions L_1, \dots, L_n and E_1, \dots, E_n with an edge between L_i and E_j of weight $|L_i \cap E_j|$. A complete matching between L and E is where each vertex in L is matched to a unique vertex in E . The total weight of the matching is the sum of the edge weights in the matching. This was solved using NetworkX's implementation of the blossom and primal-dual method (Galil, 1986).

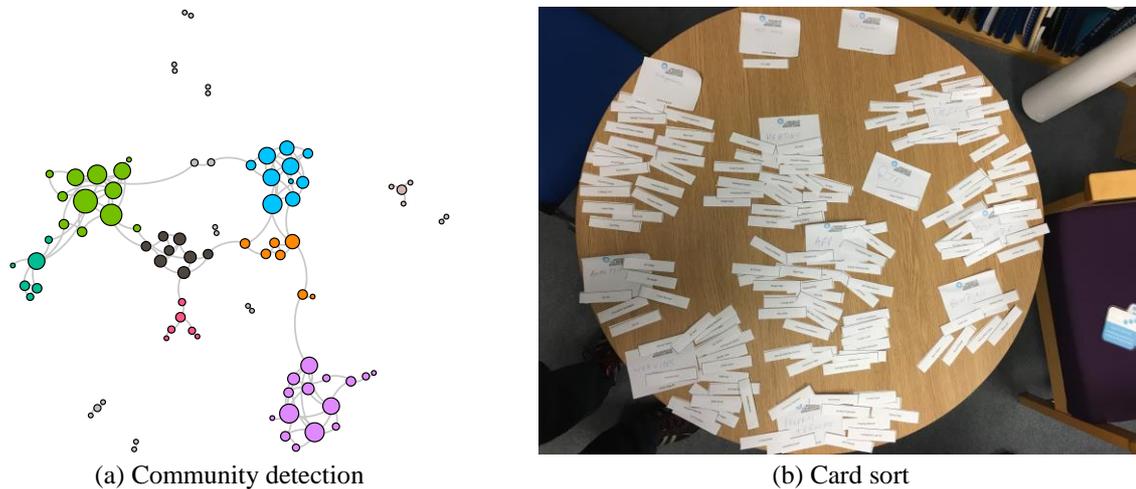


Figure 4. Graph and expert keyword partitions

Table 4 details the co-word partitioning and edit distances required to move from one partition to another. The Louvain partitioning provides the highest modularity of the graph as one would expect as it is driven by this objective function. However, the expert partitioning appears to not reflect the nature of co-occurrence with $Q = 0.254$.

Table 4. Co-word, expert and random partition comparison

Measure	L	E	
Partitions	17	9	
Q	0.824	0.254	
Comparison	L-E	R-E	L-R
Terms	87	87	87
Maximum Weight Matching	40	24	23
Minimum Edit Distance	47	63	64

Looking at the minimum edit distance, Louvain and Expert is the smallest although half of the terms would still need to be re-organised to create matching groups. Both have similar edit distances to Random partitioning, which demonstrates a systematic approach to structuring the data be it through the co-occurrence of bi-grams or expert domain understanding.

Figure 5 provides the comparison of the topics generated by Louvain, Expert and Random with the gradient highlighting the number of same bi-grams within topics and blue dots indicating the maximum edge matching between the sets of topics. It can be seen that the contribution to the edit distance has come in the form of the additional partitions produced by Louvain, which are not captured by the expert. As the partitioning algorithm is a form of hierarchical clustering, it may be that a threshold set on the objective function would produce coarser partitions that are more in line with the expert partitioning.

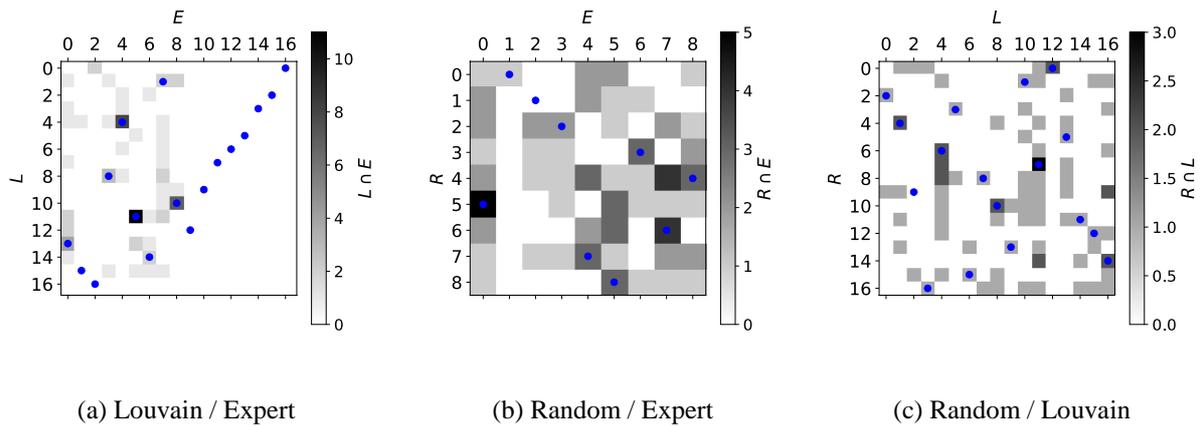


Figure 5. Bi-partite matrix and max-weight matching

5. Discussion and future work

The adaption of co-word analysis to parse technical reports through the identification of bi-grams and application of a bi-gram index distance to determine co-occurrence has shown promise in identifying elements of knowledge and partitioning of knowledge elements into topics.

The potential of using graph properties to evaluate whether a bi-gram is a knowledge element has been confirmed with T-tests showing significant difference in the properties of bi-grams determined to be knowledge elements by experts. Having proved this, one can now go further in determining the thresholds and boundaries one needs to set for a bi-gram to be automatically considered a knowledge element. In addition, opportunities exist in using supervised learning where experts label a portion of the bi-gram co-word graph and allow the algorithm to propagate the labels across the graph.

The topic identification through graph partitioning revealed a structure whose edit distance is considerably less than that of random partitioning. This demonstrates that partitioning could pose as an expert in mapping knowledge although further investigation is required to determine the nature of the differences between expert and algorithmic partitioning. To achieve this, future work is planned to perform more expert knowledge map exercises to investigate the difference in knowledge maps between experts and whether the edit distance seen in this experiment is comparable to the variance observed in expert opinion.

It is also the case that the algorithm and analysis has been performed on a subset of reports. Thus, the algorithm currently lacks all the information to construct a high-fidelity knowledge map whereas the group of experts would have additional understanding of the subject matter, which they would have brought to the term partitioning exercise. This would be mitigated in subsequent studies by increasing the sample size.

The final aspect to consider is the sensitivity of the analysis due to the index distance used to determine co-occurrence as well as the size and variety of terms used in the document set. Further work could be performed to determine how the graph topology varies with these parameters. In addition, further partitioning techniques could be explored to see which best correlates with expert defined partitions, such as multicores-periphery structures (Yan and Luo, 2019).

6. Conclusion

Design & Manufacture Knowledge Mapping continues to be a critical activity and the synthesis of maps is becoming ever more challenging as the breadth of knowledge required by organisations is increasing in order to develop products of increasing complexity. This paper has proposed and demonstrated that an adapted model of co-word analysis can complement and may even supersede the current industry standard knowledge mapping techniques. T-tests have shown that graph properties can be used to distinguish bi-grams that represent elements of knowledge and partitioning can be used to group knowledge elements into topics.

Acknowledgements

The work reported in this paper has been undertaken as part of the UKRI NCC Researcher in Residence (EP/R513556/1) and EPSRC Trans-disciplinary Engineering (EP/R013179/1) grants.

References

- Björnfort, A. and Stehn, L. (2007), "A design structural matrix approach displaying structural and assembly requirements in construction: A timber case study", *Journal of Engineering Design*, Vol. 18 No. 2, pp. 113-124.
- Blondel, V.D. et al. (2008), "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008 No. 10, p. 10008.
- Deibel, K., Anderson, R. and Anderson, R. (2005), "Using edit distance to analyze card sorts", *Expert Systems*, Vol. 22 No. 3, pp. 129-138. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0394.2005.00304.x>
- Ding, Y., Chowdhury, G.G. and Foo, S. (2001), "Bibliometric cartography of information retrieval research by using co-word analysis", *Information processing & management*, Vol. 37 No. 6, pp. 817-842.
- Dori, D. and Shpitalni, M. (2005), "Mapping knowledge about product lifecycle engineering for ontology construction via object-process methodology", *CIRP annals*, Vol. 54 No. 1, pp. 117-122.
- Galil, Z. (1986), "Efficient algorithms for finding maximum matching in graphs", *ACM Computing Surveys (CSUR)*, Vol. 18 No. 1, pp. 23-38.
- Gopsill, J. et al. (2015). "The evolution of terminology within a large distributed engineering project", In *International conference on engineering design (iced15)*.
- Gopsill, J.A. et al. (2016), "Automatic generation of design structure matrices through the evolution of product models", *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, Vol. 30 No. 4, pp. 424-445.
- Hagberg, A.A., Schult, D.A. and Swart, P.J. (2008), "Exploring network structure, dynamics, and function using NetworkX", In *Proceedings of the 7th python in science conference (scipy2008)*, pp. 11-15.
- He, Y. et al. (2019), "Mining and Representing the Concept Space of Existing Ideas for Directed Ideation", *Journal of Mechanical Design*, Vol. 141 No. 12, p. 121101.
- Jafari, M. et al. (2009), "A framework for the selection of knowledge mapping techniques", *Journal of Knowledge Management Practice*, Vol. 10, p. 1.
- Jones, D.E. et al. (2015). "Improving enterprise wide search in large engineering multinationals: A linguistic comparison of the structures of internet-search and enterprise-search queries", In *Ifip international conference on product lifecycle management*, Springer, pp. 216-226.
- Jones, S.L. et al. (2015). Subject lines as sensors: Co-word analysis of email to support the management of collaborative engineering work. In *Ds 80-6 proceedings of the 20th international conference on engineering design (iced 15) vol 6: Design methods and tools-part 2*, 27-30 July 2015, pp. 307-318.
- Le, Q. and Mikolov, T. (2014), "Distributed representations of sentences and documents", *International conference on machine learning*, pp. 1188-1196.
- Liu, Y. et al. (2014). "Chi 1994-2013: Mapping two decades of intellectual progress through co-word analysis", In *Proceedings of the 32nd annual acm conference on human factors in computing systems*, pp. 3553-3562. ACM.
- Newman, M.E.J. (2004), "Analysis of weighted networks", *Physical Review E*, Vol. 70 No. 5, p. 056131.
- Newman, M.E.J. (2006), "Modularity and community structure in networks", *Proceedings of the National Academy of Sciences*, Vol. 103 No. 23, pp. 8577-8582.
- Sarica, S., Luo, J. and Wood, K.L. (2020), "Technet: Technology semantic network based on patent data", *Expert Systems with Applications*, Vol. 142, p. 112995.
- Sarica, S. et al. (2019), "Engineering knowledge graph for keyword discovery in patent search", *Proceedings of the Design Society: International Conference on Engineering Design*, Vol. 1 No. 1, pp. 2249-2258.
- Schmidt, D.M. et al. (2015). "Identification of knowledge and processes in design projects", In *Ds 80-10 proceedings of the 20th international conference on engineering design (iced 15) vol 10: Design information and knowledge management*, 27-30 July 2015, pp. 283-292.
- Schmidt, D.M. et al. (2013a). "Interpreting knowledge maps using structural criteria", In *Ds 75-6: Proceedings of the 19th international conference on engineering design (iced13), design for harmonies, vol. 6: Design information and knowledge*, 19-22 August 2013.
- Schmidt, D.M. et al. (2013b). "Multiple-domain matrices and knowledge maps for visualizing knowledge-driven scenarios", In *Reducing risk in innovation: Proceedings of the 15th international dsm conference*, 28-30 August 2013, p. 55.
- von Saucken, C. et al. (2014). "Measures and methods for systematic knowledge management", In *Ds 77: Proceedings of the design 2014 13th international design conference*, pp. 1925-1936.
- Yan, B. and Luo, J. (2019), "Multicores-periphery structure in networks", *Network Science*, Vol. 7 No. 1, pp. 70-87.