Shared, Shaped, and Stolen: Tracing Sites of Knowledge Transfer across Creative Communities of Practice

Adam G. Emerson*
adam.emerson@mavs.uta.edu
University of Texas at Arlington
Arlington, Texas, USA

Shreyosi Endow*
shreyosi.endow@mavs.uta.edu
The University of Texas at Arlington
Arlington, Texas, USA

Cesar Torres
cearto@uta.edu
University of Texas at Arlington
Arlington, Texas, USA

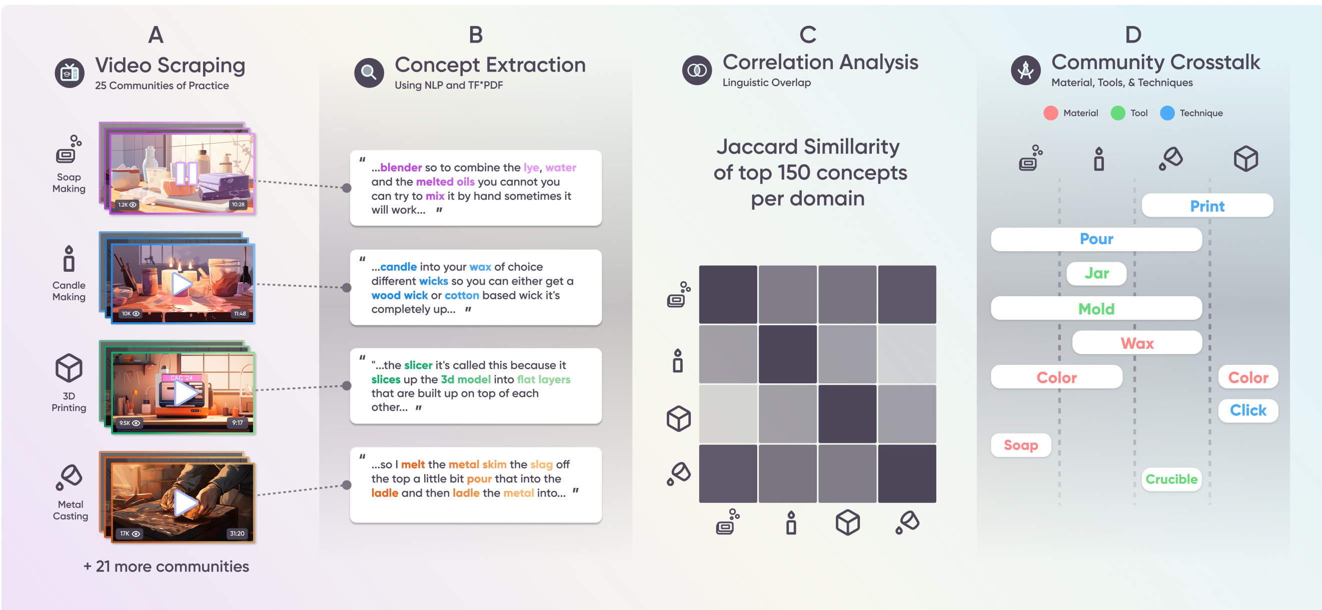

Figure 1: A) Metadata and transcripts are collected from 1.8k video tutorials across 25 creative Communities of Practice (CoPs); B) Part-of-speech tagging and TF*PDF ranking on tutorial transcripts is used to extract and highlight the most relevant concepts for each community; C) A similarity matrix with agglomerative hierarchical clustering is used to draw insight on semantic overlaps between communities; D) Using a LLM, concepts are tagged as "Material", "Tool", and "Technique", providing a more detailed look at the interplay between adjacent and disparate communities.

ABSTRACT

Within various creative domains, communities of practice (CoPs) are instrumental in fostering knowledge creation and innovation. Although each community disseminates knowledge through resources like online video tutorials, this content is often hidden behind different contexts and semantics that limits practitioners' ability to learn, borrow, and adapt knowledge from each other. To trace how knowledge disseminates across CoPs, we analyzed video transcripts across 25 communities and characterized them using Term Frequency Proportional Document Frequency (TF*PDF) to extracted materials, tools, and techniques concepts. Using a cluster heatmap visualization, we reveal material and material parallels as boundaries for *umbrella CoPs*, techniques as strong predictors of *kindred* CoPs, and outliers as emerging sites of *hybrid CoPs*. We discuss implications for the design of knowledge discovery support tools to characterize material workflows, track knowledge evolution, and develop semantic vocabularies.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Applied computing** → *Fine arts*.

KEYWORDS

natural language processing, skill acquisition, semantic web

ACM Reference Format:
Adam G. Emerson, Shreyosi Endow, and Cesar Torres. 2024. Shared, Shaped, and Stolen: Tracing Sites of Knowledge Transfer across Creative Communities of Practice. In *Creativity and Cognition (C&C '24), June 23–26, 2024,*

*Both authors contributed equally to this paper.

C&C '24, June 23–26, 2024, Chicago, IL, USA

ACM ISBN 979-8-4007-0485-7/24/06
https://doi.org/10.1145/3635636.3656199

Chicago, IL, USA. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3635636.3656199

1 INTRODUCTION

Online video tutorials have evolved beyond mere informational tools, emerging as pivotal anchors for forming vibrant online communities of practice (CoPs). These tutorial-driven communities often coalesce around specific hashtags or content niches (e.g., #resintok), creating virtual hubs of expertise and exploration that through the continual engagement and camaraderie of their members, epitomize communities of practice [24]. Readily encountered in skill-based and craft activities, CoPs develop their own repertoire of form-giving techniques which involve the selection and manipulation of materials, the application of specific tools and techniques, and the implementation of design principles to create tangible forms. These techniques are essential for translating conceptual ideas into physical realities and are often inspired from "other adjacent and parallel practices, from which lessons are learned, innovations borrowed, procedures copied" [6, 51].

Form-giving techniques are inherently interconnected in nature. Within HCI, there exists a renewed focus on developing a vocabulary around how different physical and computational materials are used in interaction design [39], which has led HCI researchers to investigate material practices in CoPs such as ceramics [20, 41], textiles [4, 9, 13, 21], and glass [38]. Despite research efforts to facilitate interdisciplinary creativity in person [17], supporting CoP *cross-talk* – a form of tacit knowledge transfer between CoPs sharing common methodologies, tools, and techniques – through digital interactions remains challenging since such communities are often insular [24]. Knowledge sharing among practitioners is often confined within community-specific hashtags and recommendation algorithm-driven filter bubbles, which selectively promotes content similar to what practitioners already engage with [43]. As a result practitioners have a limited ability to access, consolidate, or synthesize knowledge from domains outside of their expertise which can hinder potential creative possibilities in their respective fields.

Tuning into the cross-talk among CoPs can provide HCI researchers with a deeper, more nuanced understanding of the similarities and differences between how practices interact with materials and tools, broadening the interaction design opportunities for computational tools that support materials and creative practices. To understand sites for CoP cross-talk, we analyzed communities of practices leveraging the online video tutorials. These tutorials offer a view into the language, materials, tools, and techniques of each community. We adopted a mixed-method approach to analyze video transcripts to better understand where knowledge transfer was occurring. Our work contributes:

- *CoP Characterization Technique* We mined 72k concepts from 1.8k tutorial transcripts across 25 CoPs chosen from HCI-adjacent research. We leveraged LLMs to identify and enumerate key material, tool, and technique concepts mentioned in video transcripts. To characterize each CoP, Term Frequency Proportional Document Frequency (TF*PDF) was used to isolate the top 150 concepts most representative of each COP. Our findings serve to annotate the current conceptual landscape within each CoP.
- *CoP Similarity Analysis* To analyze the connections between creative communities, we created a concept-focused similarity matrix centered on materials, tools, and techniques. This matrix was reorganized using agglomerative clustering to form clusters, visually demonstrating concept overlaps among Communities of Practice (CoPs). By examining these clusters, we identified key themes that reveal the nature of knowledge exchange within these communities and introduce new terms to better describe CoP types. Our analysis highlights parallel practices and potential avenues for knowledge transfer among CoPs.

In this paper, we first describe relevant research on facilitating dialogue amongst communities of practices and concept extraction techniques. We then detail our concept enumerating technique, the resulting concept corpus and matrix, and our cluster analysis approach. Through our visualizations, themes, and design implications, our findings serve as a snapshot of creative communities that can be used to inform the design of knowledge discovery and creativity support tools that encourage cross-talk among creative communities. We discuss how our characterization technique can be used to map material workflows, track knowledge flow and evolution for supporting maker bibliometrics, and identify opportunities to further refine and develop semantic vocabularies in tacit practices.

2 RELATED WORK

Our work focuses on exploring the role of tutorials in promoting and supporting community interactions. In doing so, we situate our research within the context of ongoing endeavors to exchange knowledge within communities of practice. Furthermore, we delve into established methods that have been employed to extract meaningful and relevant concepts from unstructured texts.

2.1 Analyzing Communities of Practice

We adopt Lave and Wenger's [24] definition of communities of practice to describe "a group of people who share a common interest or activity and that belong to a social structure that reflects shared histories of learning".

HCI researchers have leveraged a plethora of design methods to analyze communities of practice, such as interviews [21, 29, 34, 41], co-design [55], and participant observation [40]. Engaging with CoP practitioners and experts through these design methods are appropriate in accessing a multitude of rich and tacit information and understanding first-hand the fundamentals of a practice.

For example, Rosner et al. [41] conducted ethnographic interviews with six ceramicists to understand their material relationships with clay and obstacles in integrating clay with technology. Devendorf et al. [9] hosted a six-week artist residency with a weaver and textile designer to understand how working with craftspeople can aid in identifying ways to approach solutions in smart textiles that counter ways in which HCI researchers/engineers approach solutions. Nicholas et al. [34] carried out semi-structured interviews with creative practitioners in diverse domains such as performance, craft, engineering, and design to identify ways in which experts engage with and manage creativity-relevant processes. Moradi et al. [31] examined silicone techniques used by DIY, soft robotics, and advance manufacturing communities of practice to inform the design of a silicone fabrication techniques. Rakib et al. [38] conducted workshops with glass artists in a glass studio to create relational ontologies that can be used to formalize and consolidate domain knowledge across CoPs.

However, creative practices are constantly evolving with the rise of new technologies and changes in economic and social demands. Analyzing a wide range of CoPs and keeping up-to-date with their characteristics and practices is challenging and costly with these existing CoP analysis methods. In this work, we present a technique to analyze CoPs using readily available video tutorials on online platforms. This technique provides CoP insights from the works of a large body of practitioners, ensuring that the analysis remains relevant without the costs associated with field research. Using tutorials also makes it possible to track how a CoP evolves through time without the need to conduct multiple studies.

2.2 Extracting Community Dialogue from Tutorial Content

Tutorials have been understood as playing an important role in facilitating dialogue among online communities of practice by bringing in an informational meta-layer that accompanies a practitioner's artifact [47], allowing for direct and contextual feedback [32], and connecting members through authorship and viewership [50].

Online tutorials often include commentary and anecdotes on the hurdles, failures, and successes encountered while making, resulting in a conversational tone not seen in other forms of technical documentation [8, 14]. Tutorial video archives and their comment sections, which often remain active long after the videos were posted, elevate the content to something closer to a "living document", relying on community engagement to keep information contemporary and accurate and to fill in information omitted by the original author [14]. Within HCI, tutorial transcript data has been used to extract coarse-grained and fine-grained events within tutorials [49].

However, these efforts in extracting information from tutorial archives and transcripts primarily improve the visibility, digestibility, and learning experience of the tasks depicted in the tutorials and not the practice as a whole. In our work, we demonstrate how tutorials also serve as a suitable medium for extracting core concepts within CoPs to provide a holistic view. Our work analyzes transcript data of existing video tutorials to identify key materials, tools, and techniques that annotate the conceptual landscape within each CoP.

2.3 Concept Extraction

Concept extraction from unstructured texts creates opportunities to reconfigure information in novel ways that improve user experiences and learning outcomes. ConceptScope [54] achieves this by leveraging ontologies, but these are static and require maintenance and intervention. In our work, we extract concepts from a dynamic source of documents (video tutorials) ensuring that the corpora remains contemporary to its domain.

Researchers have also employed probabilistic models to extract multi-word concepts or phrases. El-Kishky et al. [12] introduced the ToPMine algorithm which efficiently mined candidate phrases from corpora, outperforming other attempts at the same task. Li et al. [25] explored using vector embeddings to mine concepts using semantic context. We leverage established natural language processing techniques such as part-of-speech and named entity tagging and stop-word elimination to extract concepts from video transcript data. We then use a large language model (GPT-4) to label concepts as materials, tools, and techniques, and further verify these labels with human raters.

Concept extraction also often depends on term weighting techniques such as the popular algorithm, Term Frequency Inverse Document Frequency (TF-IDF) [46], which highlights the words best suited to identifying a single document in a corpus. We use an alternative term-weighting scheme, Term Frequency Proportional Document Frequency (TF*PDF) [3], which aids in identifying terms that are widely shared and discussed within a corpus and is often used as a tool for identifying emerging topics in a target domain [33].

3 COP CHARACTERIZATION TECHNIQUE

We detail our method for collecting data from various maker communities and motivate our design principles for organizing and quantifying the data in a way that is representative of the various CoPs.

3.1 Selecting Creative Communities

We initiated our research by conducting a review of existing SIGCHI literature constrained to the last 10 years to identify previously studied communities and practices with a significant online presence, querying on the terms "material", "maker", "DIY" and "practice". This approach was used to reflect current trends in the HCI field.

From this corpus, we identified key communities such as ceramics [14, 30, 41], 3D printing [10, 22, 48], and textiles [2, 11, 13, 21]. These communities were chosen based on their diverse range of practices, from traditional crafts like weaving to more contemporary applications such as electronic embroidery [13, 37].

In addition to the communities identified through the literature review, we also opted to include several communities with a strong online presence but less direct representation in academic research to better capture a broader spectrum of current DIY practices, including those that are thriving in online spaces but are yet underrepresented in scholarly studies. These communities, such as candle making and soap making, are characterized by their vibrant online activity and significant community engagement.

To ensure a detailed analysis of each DIY community, we subdivided CoPs into specific more granular CoPs based on an analysis of online content. For instance, the broad category of 'Textiles' was divided into 'Knitting', 'Crochet', 'Weaving', and 'Embroidery', each representing a unique subset of practitioners with distinct techniques.

The variety within these communities provides an opportunity for a broad analysis of knowledge-sharing behaviors, techniques, and the development of practices in physical environments.

Although we initially arrived at 30 communities for inclusion in our Concept Corpus, we opted to exclude 5 of them which we felt ultimately did not share the same motivations as our other CoPs. Automotive Repair, for example, while an abundant source of video tutorials, is more focused on utility and pragmatism rather than creative expression.

3.2 Harvesting Video Tutorials

We established a systematic approach to harvest video tutorials using the `Youtube-Transcript-API` [7] which allowed us to collect and process video metadata and transcripts simultaneously.

Seed Keywords. Leveraging the API search endpoint, tutorials were gathered using the search phrase "[CoP Title] tutorials" (e.g., "metal casting tutorials" or "3D printing tutorials").

Scraping Video Transcripts. Metadata was collected for $1.8k$ videos, including information such as `video ID`, `title`, `view count`, `like count`, and `duration`. Our review of videos indicates that relevancy started to taper off at $n = 100$ videos for a majority of CoPs.

Transcripts were then obtained in JSON format and organized into individual lines with `text`, and `start_time` and `end_time` (in seconds). Transcripts generated by the Youtube-Transcript-API lacked punctuation and case distinction, but these elements were not necessary for the NLP techniques used in this work, so no further modifications were made to the raw transcripts. An excerpt from a raw transcript for a soap making tutorial [28] is shown in Table 2.

Some videos had either no transcripts enabled or had transcripts consisting solely of descriptions of sound effects like "[music]" and "[applause]". In some cases, videos featured text overlays instead of audio narration. Transcripts for these videos were available only if the tutorial author had added the text overlays as closed captions explicitly. Consequently, as some of these videos were pruned, the overall count of transcripts per domain varied.

3.3 Concept Extraction and Selection Metrics

To identify community concepts, we performed part-of-speech tagging and lemmatization on each transcript using `spaCy` [19], extracting the nouns and verbs. These extracted concepts are stored within an array alongside the original transcripts, the tutorial metadata, and the search phrases in a SQL database. The database schema is shown in Table 1.

Search Terms

	searchID	int
	searchText	text
	domain	text

CoPs

	domain	text
	category	text

Transcripts

	id	int
	text	text
	concepts	text[ ]
	start	int
	end	int
	videoID	text

Videos

	videoID	text
	title	text
	description	text
	duration	int
	likes	int
	views	int
	numComments	int
	channelID	text
	domain	text
	searchID	int

- Primary Key
- Foreign Key

Table 1: Database Schema

Relevance Scoring. To highlight the most significant concepts and to enable a comparative analysis across CoPs, we establish a metric of relevance for concepts within each DIY community. One popular relevance metric, Term Frequency Inverse Document Frequency (TF-IDF) is widely used in information retrieval, and works by highlighting the uniqueness of individual terms within a corpus. This results in a tendency to identify terms that are relatively rare and specific to only a few documents (or transcripts). While the TF-IDF approach aligns well with traditional document and information retrieval tasks, it is less than ideal when the goal is to discover terms prevalent across many documents within a corpus. For example, consider Wood Carving, a concept with a high-ranking TF-IDF score might be "Dinosaur" because a single maker chose to carve a Dinosaur in their tutorial.

Term Frequency Proportional Document Frequency (TF*PDF) [3] (Equation 1-2) is recognized by researchers as an alternative term weighting scheme useful for identifying terms that best describe a domain as a whole [5, 33, 45]. It assigns higher weights to terms occurring frequently in many documents, effectively highlighting terms that are shared and discussed widely within a corpus. This technique aligns well with our goal of extracting concepts representative of the core knowledge within communities of practice. Below is the equation for TF*PDF, adapted to our task, where W_c is the weight of concept c, $F_{c,d}$ is the frequency of c in CoP d, $n_{c,d}$ is the number of video transcripts containing concept c in CoP d, N is the total number of video transcripts in CoP d, K is the total number of concepts in a given CoP, and D is the total number of CoPs.

$$W_c = \sum_{d=1}^{D=d} |F_{c,d}| \cdot exp(\frac{n_{c,d}}{N_d}) \tag{1}$$

$$|F_c| = \frac{F_c}{\sqrt{\sum_{k=1}^{K=k} {F_k}^2}} \tag{2}$$

For our data set, this results in a number within a range of [1 - ≈ 10000], with highly relevant concepts for each domain scoring in the thousands.

3.4 Extracting Concept Types

To better understand and organize domain concepts, we leveraged a large language model (LLM), GPT 4 [35] to classify concepts into three discrete categories; Material, Tool, and Technique. These categories were motivated by craft research where Makela et al. [27] identified how dialog between *environment*, *materials*, and *tools* facilitate a craftsperson's relationship with their practice [27]. Since tutorial transcripts cannot explicitly capture the tutorial authors' environment, we instead used Techniques as a category which represent material-tool relationships. The top 150 TF*PDF ranked

Excerpt from "Liquid Soap Making Tutorial – Complete Process and Easy Beginner Recipe"
...it depends on your approach i actually like to go the hot and fast method rather than the slow and cooler method with my liquid soap making so i do tend to crank up my crock pot to high and with this i will i will be um blasting these oils fairly well you will need a stick blender to make this soap i've got mine over here ready to go and really for cooking the soap the stick blender and a bit of patience is all you need i'm going to be using this trusty silicon spatula spoony thing because it works quite well so this coconut oil is nearly melted because i use fairly high heat because i don't really want to wait around all day for this soap to to come together i don't leave it i supervise it pretty much the whole way through and that you know it makes it easier if you can do it faster there are some so liquid soap making methods and recipes that take hours and hours and hours i like to do the quicker version now that my oils are well and truly all melted together and getting quite warm i'm going to add the lye solution you just pour it all in so you can see there's a lot of water in that soap but it does make it a lot easier to mix what i've found with liquid soap making is every recipe behaves quite differently so you've really just got to work with a recipe that...

Table 2: An excerpt from a video transcript with classified concepts highlighted as Material, Tool, Technique, and Other. Timestamps were removed for clearer presentation.

concepts were selected from each domain and provided to the LLM along with the domain title. The LLM was given a definition and examples for the three classes and was tasked with assigning each concept to a class. In the case that a concept was encountered which did not neatly fit into one of these predefined classes, the LLM was instructed to label it as Other. The full prompt used is available in § A.2. To validate the LLM output, two paper authors independently labeled the top 100 concepts from the Laser Cutting domain where consensus was reached on 80% of concepts. Agreed upon human generated labels were compared against the LLM output, and it was found that consensus could be reached for 89% of target terms (71% material, 87% tool, 96% technique); of the times that labels mismatched, more than half (58%) were due to the LLM labeling concepts as 'Other'. This often happened with ambiguous concepts where it was difficult to distinguish between materials, tools, and techniques. For such concepts, human raters still typically assigned one of these categories, whereas the LLM opted for "Other".We associate the lower material accuracy due to the digital design nature of lasercutting. Table 2 demonstrates how concepts included in the top 150 ranked concepts in the Soap Making domain were classified by the LLM.

3.5 Generating Concept Corpus

Following the video transcript concept enumeration method, we generated a corpus of concepts from the 25 selected creative communities. The number of concepts extracted from each CoP is provided in § A.1. Notably, the Coding CoP had the longest average video duration (85.5 min), while the Sugar Working CoP had the shortest (6.24 min). Sugar working, Coding, and Crochet had the most material, tool, and technique concepts within their top 150 concepts, respectively.

4 HIERARCHICAL CLUSTER ANALYSIS

To explore the relationships between creative CoPs, we generated a similarity matrix visualized as a heatmap to help visually identify the degree of association between different communities. Using this visualization, we identified CoP clusters which were then used to seed a qualitative analysis of concept overlap within these communities. We present a set of themes to describe the type of knowledge transfer occurring within these CoP clusters.

Matrix Generation. In this study, a 25 x 25 similarity matrix was constructed to analyze the relationships between the 25 CoPs in our corpus (Figure 2). For each CoP, we identified the top 150 concepts based on their TF*PDF ranking. The Jaccard index, also known as the similarity coefficient, was then used to compute similarity between CoP A and CoP B as follows:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

Unlike other similarity measures that might be influenced by the size of the document or the absolute frequency of terms, the Jaccard index focuses solely on the presence or absence of terms, making it a more robust and appropriate choice for our comparative analysis. By measuring the size of the intersection (common concepts) relative to the size of the union (total distinct concepts) of two sets, it provides a normalized, intuitive, and direct comparison of concept overlap and uniqueness across different CoPs making it especially effective for heatmap visualization. The index is a value between 0 and 1, with 1 indicating the communities are identical (e.g. Candle Making - Candle Making) and a value closer to 0 indicating there is little concept overlap between the communities (e.g. Sugar Working - Coding, Jaccard index - 0.07).

To make similarity clusters more salient, we adopted an unsupervised agglomerative clustering technique well-suited for hierarchical structure discovery in data. This approach starts by treating each data point as a separate cluster and iteratively merges them based on similarity measures, revealing multi-level structures that are otherwise not apparent. Since concepts were sourced a shared field (creative practices), UPGMA (Unweighted Pair Group Method with Arithmetic Mean) was used as our linkage method since it assumes a constant rate of similarity (or dissimilarity) across all pairs of elements.This assumption was further confirmed by observing similar cluster groups when using average and Ward metrics. To minimize over saturated hues on the heatmap identity (diagonal), we scaled the Jaccard index down to a maximum of 0.6 to mirror

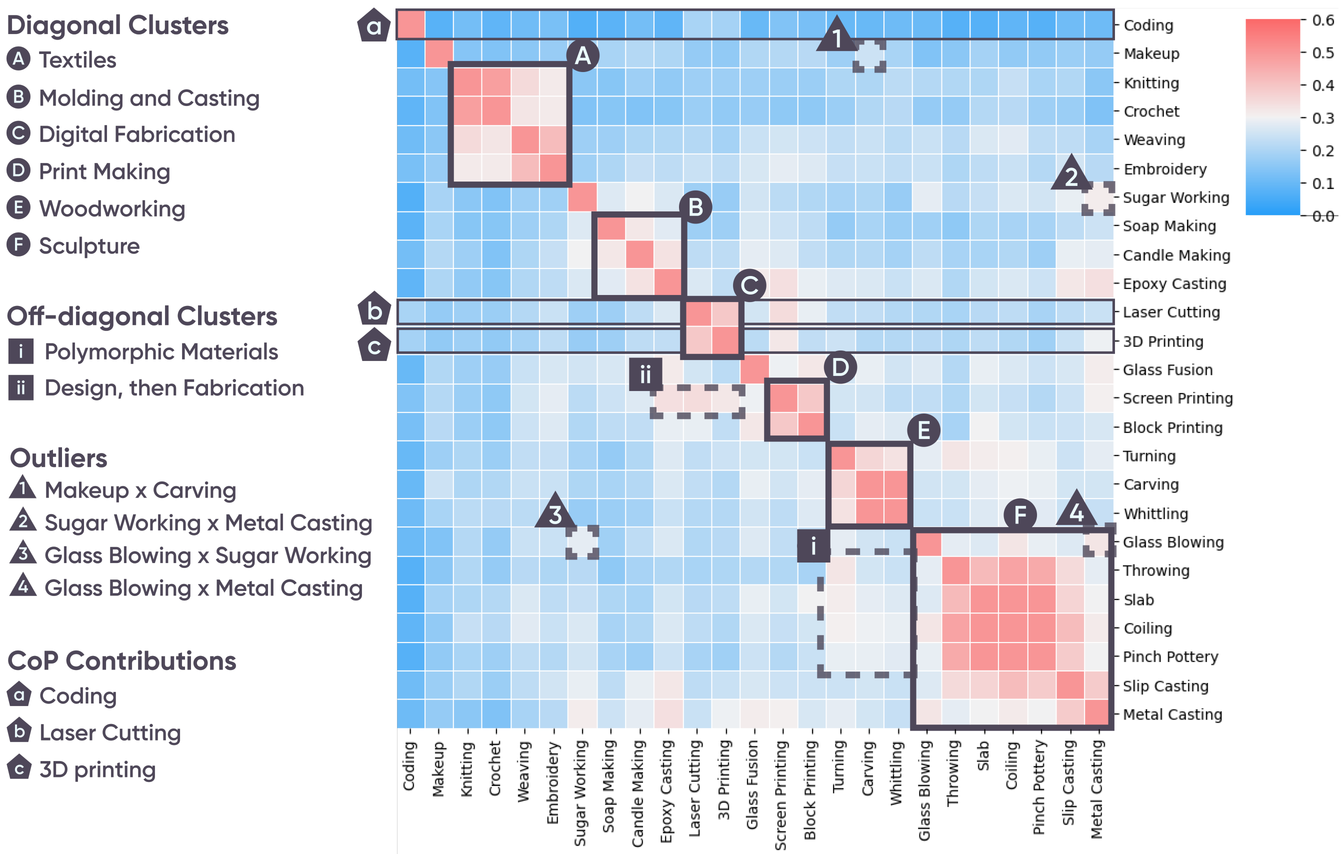

Figure 2: CoP Similarity Cluster Heatmap. A visualization of the shared concepts between 25 CoPs. The Jaccard index is used to computer similarity between CoP A and CoP B. UPGMA (Unweighted Pair Group Method with Arithmetic Mean) is used as the linkage method to perform unsupervised agglomerative clustering for hierarchical structure discovery. CoP rows and columns are arranged such that practices that share more concepts with each other are placed closer together.

the data distribution ($\mu = 0.26 \pm 0.17$) and used a diverging color palette. As a result, the CoPs rows and columns were rearranged such that practices that share more concepts with each other were placed closer together on the heatmap[1].

Cluster Analysis. To extract insights from the CoP Correlation Heatmap, we employed a systematic approach to extract meaningful insights. Our primary focus was on identifying significant visual clusters within the heatmap and concentrated on three key areas: clusters along the diagonal (identity line), those in the off-diagonal spaces, and specific patterns in the rows of the matrix. This approach allowed us to capture both the intrinsic similarities within individual communities (as shown on the diagonal) and the inter-community relationships (evidenced in the off-diagonal areas). We set a threshold for cluster consideration based on the Jaccard index, disregarding any clusters where the similarity between communities was below the mean value. This decision was made to ensure our analysis focused on the most relevant and significant relationships. From these preliminary clusters, we extracted the member CoPs and examined the top overlapping concepts. We further filtered these concepts to focus on how different types of concepts – specifically materials, tools, and techniques – intersected. To contextualize the overlaps, we traced the concepts back to their origins in the original transcripts and associated video tutorials. Two authors of the paper independently reviewed the video content and documented observations and insights specifically related to the concept overlaps.

We then applied an axial coding method. Our coding process was initially guided by deductive codes, which were directly derived from our concept typing categories: materials, tools, and techniques. This coding process was iterative and codes were refined until a high level of agreement (over 80%) was reached. These refined codes were then synthesized into broader themes which represents our interpretataion of key patterns and connections within clusters in the CoP Correlation Heatmap.

4.1 Themes

When describing relationships between two CoPs, we present the material, tool, and technique overlap distribution in the following

[1]The clustered heatmap was generated using the Seaborn library `clustermap` [52]

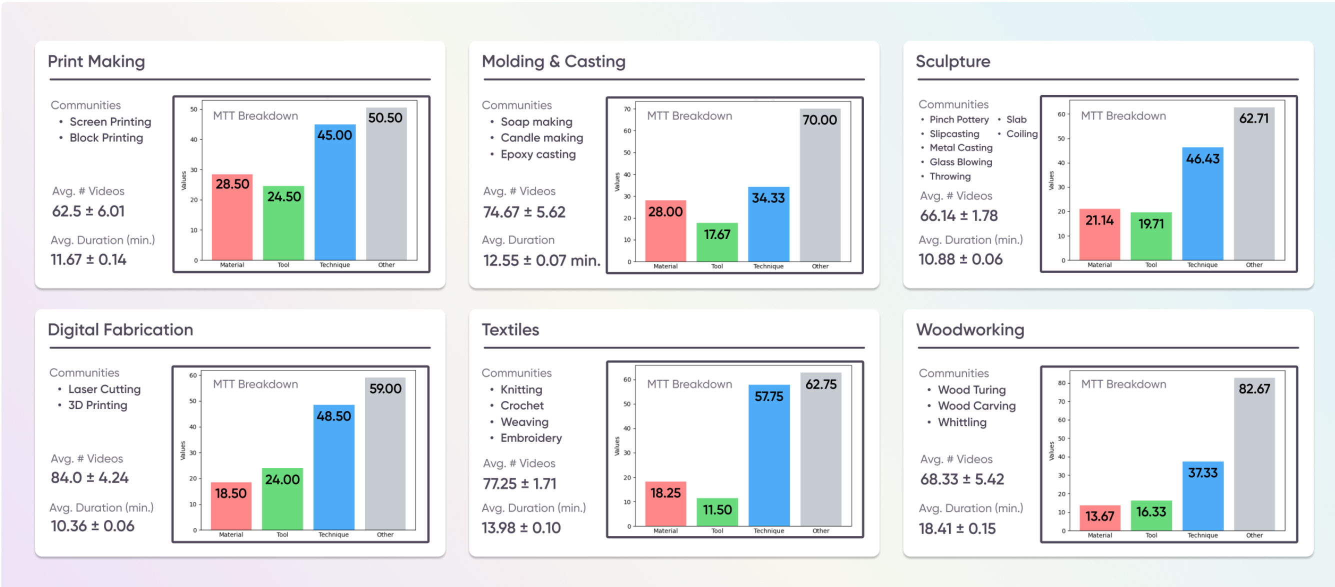

Figure 3: CoP Concept Corpus. We analyzed 25 Communities of Practice. This figure depicts data for each *Umbrella CoP*, as identified by the clusters in Fig. 2 , ordered by Material count. We characterize each CoP with the material, tool, and technique type distribution of the top 150 TF*PDF concepts within the CoP; uncategorized concepts are marked under 'Other'. All videos were sourced from YouTube using the search "[CoP Title] tutorials" Average duration is calculated with a standard error using pooled variance, and the MTT breakdown is an average of each class across the subset of domains in the cluster.

format: (# count of material concept overlap, # count of tool concept overlap, # count of technique concept overlap).

Materials and material parallels drive umbrella CoPs. The heatmap revealed six clusters which appear on the matrix diagonal and represent strong *intra-group* similarities and indicate the presence of an **Umbrella CoP**. Notably, members of these clusters had a high degree of similarity in material concepts, which indicates that sharing a common material drives community boundaries. These groups includetextiles (A), molding and casting (B), digital fabrication (C), printmaking (D), and woodworking (E), and sculpture (F) in Figure 2.

For example, the sculpture cluster (F) consists of a **family** CoP of Throwing, Slab, Coiling, Pinch Pottery, and Slip Casting – all of which have strong affinity to each other due to their focus on clay. The cluster also revealed that within the umbrella sculpture CoP, the Coiling CoP (a largely manual method where clay is rolled into long, thin ropes and then wound in layers to form a vessel) and Pinching CoP (an entirely hand-building method that forms clay from squeezing and pressing clay) are **sibling CoPs**, showing the strongest affinity especially in regards to technique (11, 10, 31). In contrast, the cluster reveals that despite sharing common materials, some CoPs show diversity in the techniques and tools they use. For example, the Coiling CoP and the Slip Casting CoP , which deals with working with liquid clay, or slip, are more akin to **cousin CoPs**, showing significantly less overlap in their technique (10, 10, 18) and capture the distinctly different methodologies between the two practices. The presence of subclusters within the sculpture umbrella (F) reveal **material parallels**, or materials that although physically dissimilar share common behavioral traits — the Metal Casting CoP, which works with liquid metal poured into molds, and the Slip Casting CoP, which deals with liquid clay, or slip, poured into molds highlights the parallel between slip and molten metal (8, 4, 24). These diagonal clusters indicate that CoP crossover forms most naturally along shared material or material-parallel boundaries.

Table 3: Communities in our dataset that did not fall under an Umbrella CoP. The initialism M/T/T/O stands for Material, Tool, Technique, and Other

CoP	Videos	Avg. Dur. (min.)	M/T/T/O
Coding	62	85.5 ± 20.16	37 / 34 / 29 / 50
Glass Fusion	77	13.96 ± 1.51	49 / 22 / 32 / 47
Makeup	84	15.18 ± 1.00	34 / 10 / 28 / 78
Sugar Working	89	6.24 ± 0.67	37 / 24 / 41 / 48

Kindred practices share similar goals and processes. Off-diagonal clusters indicate *inter-group* similarity, or what we found were **kindred CoPs** or communities with unrelated practitioners that engage in similar methods, styles, or thematic explorations. For instance, the woodworking umbrella CoP (E: Turning, Carving, Whittling) and the ceramics family (F: Throwing, Slab-forming, Coiling, and Pinching) highlight the existence of familiar techniques in two material-divergent domains as depicted in Cluster i. In wood carving, artists start with a flat piece of wood and carve out designs or structures, removing material to achieve the desired shape or pattern. Similarly, in slab making, potters roll out flat sheets of clay (slabs) and then mold, fold, or cut them to construct ceramic pieces. Both techniques involve material being removed to create and manipulated 1.5D forms, indicating subtractive making as common ground for these kindred CoPs (2, 4, 15). Similarly, the Throwing CoP and Wood Turning CoP (1, 2, 19) both work with shaping materials from spinning them on a wheel and lathe, respectively. Notably, when clay becomes leather-hard, tools used within wood turning practices are used to trim excess clay or create cavities to share the bottom of a clay form to have a "foot". In this situation, we see clay as a **polymorphic material**, able to take on the behaviors of many different types of materials.

Consequently, off-diagonal clusters can indicate the formation of a community of interest (COIs), or communities that form from different groups united by a common goal; while rich in innovation and diversity of thought, members lack mutual awareness [18]. Cluster ii shows CoPs with strong separation between their design and fabrication phases (screenprinting, laser cutting, 3D printing, and epoxy casting). Epoxy Casting and Screenprinting (6, 5, 20) show affinity via a **process parallel**, both overlapping in concepts surrounding the drying/curing of their respective materials. Laser cutting and Screenprinting show a similar process parallel (5, 10, 22) in the predominance of needing to layer and cut 1.5D materials, albeit one with a physical material (e.g., acrylic) and the other with a digital material (e.g., SVGs).

Hybrid CoPs are isolated communities. Outliers represent sparse and abnormal similarities between two CoPs that are notably missing from respective relative CoPs. We observed four distinct outliers: Makeup and Carving (1), Sugar Working and Metal Casting (2), Glass Blowing and Sugar Working (3), Glass Blowing and Metal Casting (4). These outliers indicates the presence of a more distinct crossover, or **hybrid CoP**, creating a new community with characteristics and knowledge bases from each original group.

The Makeup and Carving CoP crossover (3, 5, 9) captures the *prosthetics CoP* – makeup artists develop latex prosthetics using a variety of carving techniques to create special effects in the film industry. The prosthetics CoP shares knowledge with its parent CoPs – the artistic flair, color theory, and application techniques from makeup is complemented with structural design, shaping, and detailing techniques from carving, yet both resonate from the precision and control required of both practices. The concept overlap in these two CoPs, however, is focused on the face as a material (concepts: eye, line, nose).

The other outliers show a much more extensive material parallel, or **material nexus** with sugar, glass, and metal and thermoforming techniques that treat heat as a co-material with glass/metal/sugar. The *Sugar Working CoP*, for instance, ports over glassblowing techniques to work with sugar's thermodynamics to similarly inflate and heat sugar into transparent and edible structures.

CoP knowledge shows evolution and spread through language. Examining the row/column of the similarity matrix indicates the contributions of a singular CoP to other creative practices and reveal patterns of knowledge transfer and technique evolution across different domains. The data points to the fact that more established CoPs, such as woodworking and ceramics, have a higher degree of overlap in techniques and shared knowledge. This is likely due to the extensive time that these communities have had to establish connections and facilitate knowledge transfer. It raises questions on the timeframe required for techniques to migrate and become ingrained within different practices. Eventually techniques may lose their 'interdisciplinary' label and be simply seen as integral parts of a given practice.

The Coding CoP (row a), for instance, is largely dissimilar from the other 24 communities however it shares a small overlap with other computationally based forms of fabrication (3D Printing (row b) and Laser Cutting (row c)). Despite their shared computational basis, these domains remain semantically distinct, which may be in part due to the large amount of technical terminology and jargon associated with programming tutorials. Outside of digital fabrication, there is an indication of knowledge overlap between Coding and the textile family, specifically Weaving and Embroidery. Words like 'String' and 'Loop', common to both CoPs, could be written off as coincidental overlaps, or they might reflect the historical connection between the domains, dating back to Ada Lovelace, the Jacquard Loom, or the Core Memory Project [42]. Alternatively, it is possible that this overlap could signal the rise of a new interdisciplinary community centered around e-textiles or other forms of innovative fabrication merging digital and textile techniques.

5 IMPLICATIONS FOR DESIGN

We leveraged our findings from the cluster analysis to discuss design implications and future research trajectories for knowledge discovery and creativity support tools.

5.1 Characterizing Practices

In this work, we leveraged metrics such as material, tool, and technique overlap to characterize different CoPs. Analyzing the clusters depicted in the similarity matrix revealed that CoP crossover forms most naturally along shared material or material-parallel boundaries. This finding has practical applications in educational contexts, such as the Legitimate Peripheral Participation (LPP) framework, which facilitates the integration of newcomers into a CoP [23]. For instance, Fiesler et al. [16] observed how members of a fanfiction CoP acquired computational and design skills through LPP processes. Our cluster analysis can enhance such educational strategies by pinpointing the most effective intersections between different practices. An example from our research shows that for someone experienced in soap-making who wishes to enter a ceramics CoP, learning slip casting might be the most beneficial starting point due to its significant overlap in techniques with soap making.

Moreover, while our current focus has been on the tangible aspects of a CoP, such as materials, tools, and techniques, future research could benefit from exploring concepts that delve deeper into the practitioners themselves. Shifting the lens to the practitioners could reveal more about the implicit, tacit aspects of a practice. For example, Wood et al. [53] suggested a method of 'destructive analysis' to extract and document the underlying principles used by experts before these principles become second nature to them. Endow et al. [15] demonstrated how analyzing the way users externalize material interactions as written descriptions could uncover similarities in user mental models of materials. These approaches could serve to expand the data used to understand material boundaries and material parallels while providing valuable insights for designing more effective learning and integration strategies within these communities.

5.2 Developing Semantically Rich Vocabularies

In this work, we leveraged part of speech and named entity tagging to extract concepts which we then categorized into materials, tools, and techniques. At its core, the success of this technique hinges on the presence of semantically rich vocabularies within CoPs. Semantically rich vocabularies include metaphoric, tacit, and experiential ways of describing materials, tools, and techniques. For instance, makeup techniques are rife with metaphorical names rooted in elemental references ("a smokey eye", "dewy look"), cultural references ("French girl look"), and even food references ("cakey foundation", "glazed lips"). In contrast, domains such as woodworking, have a more constrained semantic landscape, leveraging more direct and functional descriptors ("carve", "whittle", "polish"). This is reflected in the number of concepts present in each domain, as seen in Table 4. Although it would be intuitive to assume that semantically richer CoPs naturally facilitate better cross-domain dialogue due to the ease in finding overlaps and establishing common ground, this is not necessarily borne out in our data. Makeup has the third greatest number of concepts in the corpus, but shares very little overlap with other CoPs. However, despite the limited overlap, our analysis technique still captures some commonality between hybrid CoPs such as makeup and carving. We also observe sparse but interesting material overlaps between seemingly disparate domains such as weaving and coding which may signal the presence of new CoPs such as e-textiles. Here, the common material, "string", encompasses both tangible strings in weaving and data types in coding, and beckons a broadening of our definition of materials in digital spaces. As HCI broadens its material horizons [39], the need for *material literacy* intensifies and extracting and formalizing semantically rich vocabularies for creative practices become paramount for a future where material literacy is not just a luxury but a necessity.

5.3 Establishing CoP Cross-talk

Our findings indicate a nuanced form of interaction among CoPs, which we term as "cross-talk". By "cross-talk", we refer not to direct dialogues or explicit exchanges between communities but to an implicit or tacit transfer of knowledge that occurs through shared methodologies, tools, and techniques. This form of cross-talk suggests that while kindred CoPs may not engage in direct communication, they leverage similar tools and techniques that translate into overlapping methodologies or goals. These similarities can inadvertently facilitate the transfer of knowledge and practices, even in the absence of direct interaction. We find that kindred CoPs lack the stronger ties that sibling or cousin CoPs share around a common material, but instead leverage similar tools and techniques that translate into overlapping methodologies or goals. These practices are often complementary, and may mutually benefit from exposure to the techniques, yet lack the exposure or awareness of other kindred CoPs. Knowledge discovery and creativity support tools can benefit from creating opportunities for kindred CoP connections to occur. The design of creative spaces has been proposed as one vehicle to leverage different practices through physical collocation and navigating the social factors the influence agency and knowledge distribution [1]. Such spaces, as identified in master ceramic studios [31], are created through lifelong relationships with materials that accumulate artifacts, tools, techniques from interactions with other practitioners. In this way, identifying kindred practices could inform which practices to bring together that could mutually benefit from the diversity of tools and methods of their respective practices and their way of thinking through a material. Shared resources or spaces can act as a catalyst for interaction and crossover between differing CoPs.

5.4 Tracking Creative Practices

By sourcing data from online video tutorials, our approach presents a low-cost and replicable alternative to traditional CoP ethnographic methods and offers the ability to gauge hidden or less apparent connections between a wider breadth of practices. Outliers in the cluster analysis, for instance, were identified as hybrid CoPs and were notably isolated from other communities in the concepts they shared. Our visualizations serve as a snapshot of creative practices to archive for future researchers and practitioners. Our method can offer significant insights in tracking the progress and impact HCI research on creative practices serving as a "maker bibliometric" with specific value to initiatives aimed at evaluating broadening participation efforts in computing or understanding hybridization efforts, such as in e-textiles. Like ethnographic methods, our method also benefits from leveraging real-world data which are invaluable as they reflect organic interactions and developments within CoPs. This approach allows for a more authentic and dynamic understanding of how CoPs evolve and influence each other, which can build onto efforts to track HCI efforts in the wild [26].

6 LIMITATIONS

Online video tutorials on platforms such as YouTube have diverse authors, whose credibility is often measured by metrics such as their subscriber counts, the average number of likes and comments on their videos, and how often they show up in relevant searchers. Social metadata, however, can potentially inadvertently sideline newer authors and fix the spotlight on more established authors. This phenomena mirrors patterns observed in other open platforms, like Wikipedia, where a limited cohort of editors contribute to a substantial proportion of the content, emphasizing the concentration of influential creators in open-source environments [36]. When harvesting video tutorials, Youtube's search endpoint returned the most relevant tutorials to our search phrases; future iterations can seed tutorials by alternate metrics such as date posted to investigate the emergence of newer concepts and how CoPs evolve with time.

Additionally, with the evolution of digital platforms, there has been a noticeable shift in tutorial styles from traditional "how-to" videos to brief, eye-catching, and high-production quality "shorts". These newer styles of tutorials often eschew voiceovers, leaning instead on striking visuals and distinct stylistic elements. Given their unique nature, transcript analysis may not be always suited for extracting concepts, suggesting the need for newer techniques tailored to this evolving content landscape. Furthermore, social metadata such as likes and views may not be reflective of the CoP structure for these types of videos as they may be purely due to their production value; instead, analyzing comments can shed insight on knowledge dissemination in the community.

Concept Extraction. In this work, we leveraged part-of-speech tagging and lemmatization to extract concepts from the transcripts, resulting in the concepts being single words. Tools and techniques are often combinations of multiple words that if not captured together as one concept, could lose some of its context. Future work benefits from investigating multi-word concept extraction algorithms such as ToPMine [12] and AutoPhrase [44] to determine if they yield more niche and richer concepts compared to our baseline technique.

Concept Characterization. We leveraged GPT-4 to classify concepts into materials, tools, and techniques. The model often classified ambiguous concepts as "Other" even though a human rater was able to classify the concept as one of material, tool, or technique. One strategy to reduce the number of "Other" concepts could be to integrate parts of the raw transcript consisting the concept into the GPT prompt itself. This could provide additional context to reduce ambiguity and aid the LLM in making a concrete characterization. Multi-word concepts can also be used to address this limitation. For example, in the Knitting domain, "Garter Stitch" is a technique, however, separating the concept into "Garter" and "Stitch" creates a possibility for the LLM to classify garter as a tool and stitch as a technique.

7 CONCLUSION

In this study, we analyzed video tutorial transcripts to characterize creative communities of practice (CoPs). Utilizing traditional NLP techniques, we extracted key concepts from 25 CoPs and applied agglomerative clustering for similarity analysis. A large language model aided in categorizing these concepts as materials, tools, or techniques, enhancing our understanding of community overlaps. Our observations suggest that a CoP’s diverse range of tools and techniques can indicate its maturity, reflecting the depth of its subpractices and the significance of time and collective experience in its growth. Furthermore, we discovered that analyzing video transcript concepts is a cost-effective method for gaining insights into the dynamics of CoPs and the impact of HCI on digital fabrication and computational creativity. Material parallels and overlaps in tools and techniques across various CoPs revealed underlying connections between different practices, highlighting opportunities for interdisciplinary collaboration. These findings offer important design implications for CoPs, especially in fostering the evolution and interplay of materials, tools, and techniques, guiding the development of more inclusive and innovative creative spaces that nurture interdisciplinary collaboration.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their insightful feedback and the content creators whose videos were instrumental in enabling this research. This work was supported by the NSF REU Site on Hybrid Design and Fabrication (CNS-2150321).

REFERENCES

[1] Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2019. Exploring and understanding the role of workshop environments in personal fabrication processes. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 2 (2019), 1–43.

[2] Leah Buechley, Mike Eisenberg, Jaime Catchen, and Ali Crockett. 2008. The LilyPad Arduino: using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 423–432.

[3] Khoo Khyou Bun and M. Ishizuka. 2001. Emerging Topic Tracking System. In *Proceedings Third International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems. WECWIS 2001.* 2–11. https://doi.org/10.1109/WECWIS.2001.933900

[4] Alice Buso, Holly McQuillan, Milou Voorwinden, and Elvin Karana. 2023. Weaving Textile-form Interfaces: A Material-Driven Design Journey. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference.* 608–622.

[5] Kuan-Yu Chen, Luesak Luesukprasert, and Seng-cho T. Chou. 2007. Hot Topic Extraction Based on Timeline Analysis and Multidimensional Sentence Modeling. *IEEE Transactions on Knowledge and Data Engineering* 19, 8 (2007), 1016–1025. https://doi.org/10.1109/TKDE.2007.1040

[6] Matthew A. Dalton, Audrey Desjardins, and Ron Wakkary. 2014. From DIY tutorials to DIY recipes. In *CHI ’14 Extended Abstracts on Human Factors in Computing Systems (CHI EA ’14).* Association for Computing Machinery, New York, NY, USA, 1405–1410. https://doi.org/10.1145/2559206.2581238

[7] Jonas Depoix. 2023. youtube-transcript-api. https://github.com/jdepoix/youtube-transcript-api.

[8] Derek Van Ittersum. 2014. Craft and Narrative in DIY Instructions. *Technical Communication Quarterly* (2014).

[9] Laura Devendorf, Katya Arquilla, Sandra Wirtanen, Allison Anderson, and Steven Frost. 2020. Craftspeople as technical collaborators: Lessons learned through an experimental weaving residency. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13.

[10] Laura Devendorf, Abigail De Kosnik, Kate Mattingly, and Kimiko Ryokai. 2016. Probing the potential of post-anthropocentric 3D printing. In *Proceedings of the 2016 ACM conference on designing interactive systems.* 170–181.

[11] Laura Devendorf and Chad Di Lauro. 2019. Adapting double weaving and yarn plying techniques for smart textiles applications. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction.* 77–85.

[12] Ahmed El-Kishky, Yanglei Song, Chi Wang, Clare R. Voss, and Jiawei Han. 2014. Scalable Topical Phrase Mining from Text Corpora. *Proc. VLDB Endow.* 8, 3 (nov 2014), 305–316. https://doi.org/10.14778/2735508.2735519

[13] Shreyosi Endow, Mohammad Abu Nasir Rakib, Anvay Srivastava, Sara Rastegarpouyani, and Cesar Torres. 2022. Embr: A Creative Framework for Hand Embroidered Liquid Crystal Textile Displays. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 1–14.

[14] Shreyosi Endow and Cesar Torres. 2021. “I’m Better Off on My Own”: Understanding How a Tutorial’s Medium Affects Physical Skill Development. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS ’21).* Association for Computing Machinery, New York, NY, USA, 1313–1323. https://doi.org/10.1145/3461778.3462066

[15] Shreyosi Endow and Cesar Torres. 2023. Tacit Descriptions: Uncovering Ambiguity in Crowdsourced Descriptions of Motions and Materials. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference.* 2522–2536.

[16] Casey Fiesler, Shannon Morrison, R Benjamin Shapiro, and Amy S Bruckman. 2017. Growing their own: Legitimate peripheral participation for computational learning in an online fandom community. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing.* 1375–1386.

[17] Sisse Finken, Alma Leora Culén, and Andrea Gasparini. 2014. Nurturing creativity: assemblages in HCI design practices. (2014).

[18] Gerhard Fischer, Elisa Giaccardi, Hal Eden, Masanori Sugimoto, and Yunwen Ye. 2005. Beyond binary choices: Integrating individual and social creativity. *International journal of human-computer studies* 63, 4-5 (2005), 482–512.

[19] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. (2020). https://doi.org/10.5281/zenodo.1212303

[20] Melody Horn, Amy Traylor, and Leah Buechley. 2022. Slabforge: Design Software for Slab-Based Ceramics. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 1–12.

[21] Lee Jones and Sara Nabil. 2022. Goldwork embroidery: Interviews with practitioners on working with metal threads and opportunities for e-textile hybrid crafts. In *Proceedings of the 14th Conference on Creativity and Cognition.* 364–379.

[22] Jeeeun Kim, Clement Zheng, Haruki Takahashi, Mark D Gross, Daniel Ashbrook, and Tom Yeh. 2018. Compositional 3D printing: expanding & supporting workflows towards continuous fabrication. In *Proceedings of the 2nd Annual ACM Symposium on Computational Fabrication.* 1–10.

[23] Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation.* Cambridge university press.

[24] Jean Lave and Etienne Wenger. 2004. Communities of practice. *Learning, meaning and identity* (2004).

[25] Keqian Li, Hanwen Zha, Yu Su, and Xifeng Yan. 2018. Concept Mining via Embedding. In *2018 IEEE International Conference on Data Mining (ICDM)*. 267–276. https://doi.org/10.1109/ICDM.2018.00042

[26] Emily Lovell, Leah Buechley, and James Davis. 2023. LilyTiny in the Wild: Studying the Adoption of a Low-Cost Sewable Microcontroller for Computing Education. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 282–293.

[27] Maarit Mäkelä and Bilge Merve Aktaş. 2022. In dialogue with the environment: The environment, creativity, materials and making. *Craft Research* 13, 1 (2022), 9–34.

[28] Elly's Everyday Soap Making. [n.d.]. *Liquid Soap Making Tutorial – Complete Process and Easy Beginner Recipe.* Youtube. https://www.youtube.com/watch?v=_WPIMsLk1P8

[29] Nolwenn Maudet, Ghita Jalal, Philip Tchernavskij, Michel Beaudouin-Lafon, and Wendy E Mackay. 2017. Beyond grids: Interactive graphical substrates to structure digital layout. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5053–5064.

[30] Hedieh Moradi, Long N Nguyen, Quyen-Anh Valentina Nguyen, and Cesar Torres. 2022. Glaze Epochs: Understanding Lifelong Material Relationships within Ceramics Studios. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Daejeon, Republic of Korea) *(TEI '22)*. Association for Computing Machinery, New York, NY, USA, Article 4, 13 pages. https://doi.org/10.1145/3490149.3501310

[31] Hedieh Moradi and César Torres. 2020. Siloseam: A Morphogenetic Workflow for the Design and Fabrication of Inflatable Silicone Bladders. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 1995–2006.

[32] Cuong Nguyen and Feng Liu. 2015. Making software tutorial video responsive. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1565–1568.

[33] Khanh-Ly Nguyen, Byung-Joo Shin, and Seong Joon Yoo. 2016. Hot topic detection and technology trend tracking for patents utilizing term frequency and proportional document frequency and semantic information. In *2016 International Conference on Big Data and Smart Computing (BigComp)*. 223–230. https://doi.org/10.1109/BIGCOMP.2016.7425917

[34] Molly Jane Nicholas, Sarah Sterman, and Eric Paulos. 2022. Creative and Motivational Strategies Used by Expert Creative Practitioners. In *Proceedings of the 14th Conference on Creativity and Cognition*. 323–335.

[35] OpenAI. 2023. GPT-4: Generative Pre-trained Transformer 4. https://openai.com/. Accessed: January 29 2024.

[36] Felipe Ortega, Jesus M Gonzalez-Barahona, and Gregorio Robles. 2008. On the inequality of contributions to Wikipedia. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*. IEEE, 304–304.

[37] Ernest Rehmatulla Post, Maggie Orth, Peter R Russo, and Neil Gershenfeld. 2000. E-broidery: Design and fabrication of textile-based computing. *IBM Systems journal* 39, 3.4 (2000), 840–860.

[38] Mohammad Abu Nasir Rakib, Jeremy Scidmore, Justin Ginsberg, and Cesar Torres. 2023. Thermoplastic Kilnforms: Extending Glass Kilnforming Techniques to Thermoplastic Materials using Ontology-Driven Design. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 263–281.

[39] Erica Robles and Mikael Wiberg. 2010. Texturing the" material turn" in interaction design. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*. 137–144.

[40] Daniela K Rosner. 2012. The material practices of collaboration. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 1155–1164.

[41] Daniela K Rosner, Miwa Ikemiya, and Tim Regan. 2015. Resisting alignment: Code and clay. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. 181–188.

[42] Daniela K Rosner, Samantha Shorey, Brock R Craft, and Helen Remick. 2018. Making core memory: Design inquiry into gendered legacies of engineering and craftwork. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[43] Camille Roth, Camille Roth, Antoine Mazieres, Antoine Mazières, Antoine Mazières, Telmo Menezes, and Telmo Menezes. 2020. Tubes and bubbles topological confinement of YouTube recommendations. *PLOS ONE* (2020). https://doi.org/10.1371/journal.pone.0231703

[44] Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R Voss, and Jiawei Han. 2018. Automated phrase mining from massive text corpora. *IEEE Transactions on Knowledge and Data Engineering* 30, 10 (2018), 1825–1837.

[45] Nikolay Shilov and Nikolay Teslia. 2020. Comparing Statistical Measures for Discovering Emerging Terms in Scopus Publications in the Area of Decision Support in Smart City. In *2020 27th Conference of Open Innovations Association (FRUCT)*. 1–8. https://doi.org/10.23919/FRUCT49677.2020.9211000

[46] Karen Sparck Jones. 1988. *A Statistical Interpretation of Term Specificity and Its Application in Retrieval.* Taylor Graham Publishing, GBR, 132–142.

[47] Stacey Kuznetsov and Eric Paulos. 2010. Rise of the expert amateur: DIY projects, communities, and cultures. (2010).

[48] Ryo Suzuki, Junichi Yamaoka, Daniel Leithinger, Tom Yeh, Mark D Gross, Yoshihiro Kawahara, and Yasuaki Kakehi. 2018. Dynablock: Dynamic 3d printing for instant and reconstructable shape formation. In *Proceedings of the 31st annual ACM symposium on user interface software and technology*. 99–111.

[49] Anh Truong, Peggy Chi, David Salesin, Irfan Essa, and Maneesh Agrawala. 2021. Automatic generation of two-level hierarchical tutorials from instructional makeup videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[50] Ron Wakkary, Markus Lorenz Schilling, Matthew A. Dalton, Sabrina Hauser, Audrey Desjardins, Xiao Zhang, and Henry W.J. Lin. 2015. Tutorial Authorship and Hybrid Designers: The Joy (and Frustration) of DIY Tutorials. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 609–618. https://doi.org/10.1145/2702123.2702550

[51] Alan Warde. 2005. Consumption and theories of practice. *Journal of consumer culture* 5, 2 (2005), 131–153.

[52] Michael L. Waskom. 2020. Seaborn: statistical data visualization - clustermap. https://python-charts.com/correlation/clustermap-seaborn/. Accessed: January 29 2024.

[53] Nicola Wood, Chris Rust, and Grace Horne. 2009. A tacit understanding: The designer's role in capturing and passing on the skilled knowledge of master craftsmen. *International Journal of Design* 3, 3 (2009).

[54] Xiaoyu Zhang, Senthil Chandrasegaran, and Kwan-Liu Ma. 2021. ConceptScope: Organizing and Visualizing Knowledge in Documents Based on Domain Ontology. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 19, 13 pages. https://doi.org/10.1145/3411764.3445396

[55] Clement Zheng and Michael Nitsche. 2017. Combining practices in craft and design. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. 331–340.

A SUPPLEMENTAL DATA

A.1 Concept Corpus Table

Community of Practice (CoP)	# Videos	# Concepts	Avg. Duration ± STE (min)	Mat. / Tool / Tech. / Other
3D Printing	90	3729	9.28 ± 0.83	20 / 23 / 45 / 62
Block Printing	54	2389	9.98 ± 1.43	35 / 22 / 44 / 49
Coiling	72	2294	11.84 ± 1.03	14 / 18 / 49 / 69
Crochet	79	1623	11.34 ± 1.68	07 / 05 / 77 / 61
Coding	62	5268	85.5 ± 20.16	37 / 34 / 29 / 50
Carving	57	2722	18.77 ± 2.22	13 / 15 / 40 / 82
Candle Making	80	2994	10.51 ± 1.02	30 / 26 / 34 / 60
Embroidery	74	2289	10.5 ± 1.38	27 / 13 / 55 / 55
Epoxy Casting	61	2518	9.71 ± 1.05	26 / 14 / 37 / 73
Glass Blowing	62	2312	7.59 ± 1.23	19 / 25 / 50 / 56
Glass Fusion	77	3199	13.96 ± 1.51	49 / 22 / 32 / 47
Knitting	82	2785	20.08 ± 2.33	18 / 11 / 53 / 68
Laser Cutting	78	3385	11.6 ± 0.79	17 / 25 / 52 / 56
Makeup	84	3565	15.18 ± 1	34 / 10 / 28 / 78
Metal Casting	69	3642	11.47 ± 0.76	24 / 19 / 38 / 69
Pinch Pottery	69	2445	12.03 ± 1.39	21 / 15 / 50 / 64
Sugar Working	89	2301	6.24 ± 0.67	37 / 24 / 41 / 48
Slip Casting	57	2310	8.4 ± 1.16	23 / 15 / 44 / 68
Slab Building	66	2296	13.34 ± 1.64	25 / 27 / 42 / 56
Screen Printing	71	3069	12.95 ± 1.7	25 / 27 / 46 / 52
Soap Making	83	3429	16.66 ± 0.94	28 / 13 / 32 / 77
Throwing	68	2547	10.76 ± 1.35	22 / 19 / 52 / 57
Weaving	74	2548	13.53 ± 1.36	20 / 17 / 46 / 67
Whittling	80	2893	19.52 ± 2.03	18 / 10 / 29 / 93
Wood Turning	68	3334	16.81 ± 2.33	10 / 24 / 43 / 73
TOTAL	**1806**	**71886**		

Table 4: A breakdown of the video data from 25 CoPs. The top 150 TF*PDF ranking terms from each CoP were classified as Material, Tool, Technique or Other by an LLM.

A.2 Concept Classification Prompt

Material, Tool, and Technique Prompt - GPT-4

You are a craft-based information assistant. Your job is to classify concepts into one of 3 categories using the definitions provided for "Material," "Tool," and "Technique."

Please follow these rules:

(1) Create an array in brackets [ ] as your response.
(2) Provide concise responses, consisting of one word for each item in the array.
(3) For each item create a javascript object with the keys "domain", "concept" and "class", where class can be "material," "technique", or "tool". Add this object to the array.
(4) If you cannot confidently determine the classification, reply with "n/a.". Do this only as a last resort.
(5) Ensure that your response array conforms to json syntax. For example, each object should be separated by a comma, and the last object should not have a comma after it. Object keys should be wrapped in double quotes, and values should be wrapped in double quotes if they are strings.

Category Definitions:

- Material: A substance or element used in crafting or a word commonly associated with descriptions of materials.
- Tool: An instrument or device used in crafting.
- Technique: A method or skill used in crafting.

Examples Concepts for Each Category:

- Material examples: wood, metal, fabric, clay, yarn, resin, soft, hard, smooth, rough, shiny, dull
- Tool examples: hammer, needle, paintbrush, hook, blade, wheel
- Technique examples: carve, stitch, paint, throw, add, pour

Here is an example prompt and response:

Example Prompt:

```
[{"category": "Casting and Molding", "domain": "Candle Making", "concept": "pour"},
{"category": "Casting and Molding", "domain": "Candle Making", "concept": "wax"},
{"category": "Casting and Molding", "domain": "Candle Making", "concept": "burn"},
{"category": "Casting and Molding", "domain": "Candle Making", "concept": "knife"},
{"category": "Casting and Molding", "domain": "Candle Making", "concept": "football"}]
```

Example Response:

```
[{"domain": "candle making", "concept": "pour", "class": "technique"},
{"domain": "candle making", "concept": "wax", "class": "material"},
{"domain": "candle making", "concept": "burn", "class": "technique"},
{"domain": "candle making", "concept": "knife", "class": "tool"},
{"domain": "candle making", "concept": "football", "class": "other"}]
```

Real Prompt: <data>

Table 5: The prompt provided to GPT-4 to categorize the extracted concepts into three categories: "Material", "Tool" and "Technique", where <data> is the data provided to be categorized.