

Network Mathematics for Beginners

Valeria de Paiva

(joint work with Fong, Collard, Subrahmanian, Horowitz, Gao,
Kovalev, and Moss)
Berkeley Seminar

October, 2023

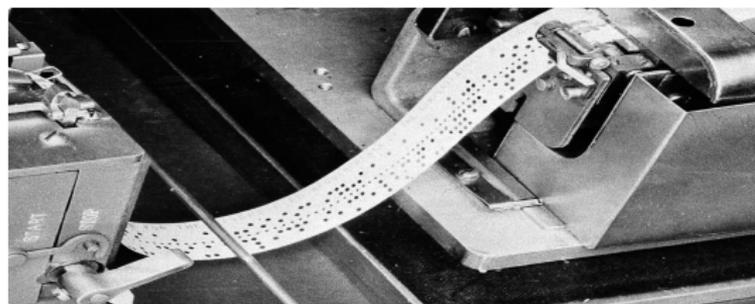


We shape technology for public benefit by advancing sciences of
connection and integration.

Our goal is a world where the systems that surround us benefit us all.

- Network Mathematics
- Research Prototypes:
 - Parmesan
<http://www.jacobcollard.com/parmesan2/>
 - MathAnnotator https://gaoq111.github.io/math_concept_annotation/
 - MathGloss <https://mathgloss.github.io/MathGloss/database>
- What's Next?

Networked Mathematics



How do we encode, connect, and share knowledge in the 21st Century?

- Make math **accessible** to everyone: students, mathematicians, scientists, interested laypeople, etc
- Making use of incredibly fast improving NLP tools
- Extract math concepts from texts & link to an ontology
- Math concepts build a Math **Knowledge Graph**, also useful for formalization
- Next year, Hausdorff Institute of Mathematics connecting our Math NLP to formalized math?

When you read some mathematics you're not familiar with...

We define the notion of a **torsor** for an **inverse semigroup**, which is based on **semigroup actions**, and prove that this is precisely the structure classified by the **topos** associated with an **inverse semigroup**. Unlike in the **group** case, not all **set-theoretic torsors** are **isomorphic**: we shall give a **complete description of the category of torsors** ...

Can we create a personal *index* of things to check automatically and on the fly?

Extracting mathematical concepts

- NLP tools are improving incredibly **fast** [but loads of hype]
- Want extracted mathematical concepts to be linked to a reasonable math ontology **WikiData**
- What linguists have for “extracting math concepts”?
- Terminology extraction (TE) for identifying fundamental mathematical concepts could be considered Named Entity Recognition (NER), keyword extraction (KE) or even definition extraction (DE)
- No ATE system dedicated to mathematics

Generic ‘Automatic Term Extractors’ (ATEs) don’t seem to work well for mathematical text (Parmesan 1, "Extracting Mathematical Concepts from Text", arxiv 2208.13830)

Baselines and Gold Standards

- Checking baselines is HARD, as they do NOT exist.
- No human annotated/golden standard datasets of math concepts?
- there was a 2017 Competition on scientific concepts. one small golden dataset, later expanded to SciERC corpus. little transfer?
- Results with *silversets* (Collard et al 2022) very low
- Video <https://youtu.be/-ZhZjMn1Zpk>

Parmesan: mathematical concept extraction for education

- a different tackle: let users ask for concepts! in <https://arxiv.org/abs/2307.06699>.
- Parmesan does entity linking and collocation search for CT, from a human learner perspective
- uses LaTeXML converter (NIST) and spaCy
- Term Extractors used:
DyGIE++ (SOTA 2019),
TextRank, mwetoolkit,
SpERT.PL(2021) and PL-Marker (2022)
- Last two recent neural term extractors, trained for science.
little transfer, see tables.
- Definition extraction using linguistic tools

Parmesan demo:

<http://www.jacobcollard.com/parmesan2/>

PARsing Math Entities Search And Navigation (Parmesan)

Parmesan 0.2

Knowledge Base

This is our best guess as to your search term's value in external knowledge bases, such as [Wikidata](#) and [nLab](#).

[Monoidal Category](#)

Wikidata

category admitting tensor products

[Monoidal Category](#)

nLab

monoidal categories

Corpora

Your search for "monoidal category" produced **1063** total results. These results are examples of use in key category theory datasets.

nLab (displaying 10 of 995 results) 

TAC (displaying 10 of 68 results) 

Generative AI for Math concepts?

- Make math more **accessible** to everyone:
- Make sure concepts extracted are correct and as complete as possible
- a kind of **WordNet** for mathematics
- Can be used to produce index for monographs, books, self-study more easily
- Understand opportunities and limitations of large language models – Old English similar problem
- Help start building a **Math Knowledge Graph**
- (eventually) Mathematical Natural Language Inference (MathNLI)

- Many mathematical terms are NOT math concepts:
characterization, conjecture, consequence, counter-example, paper, etc
- Traditional issues of terminologies:
 - plurals or not? (*presheaves or presheaf?*),
 - adjectives without a noun? (*2-categorical?*),
 - adjective with a common English noun? (*2-categorical case*)
 - adjective is related to noun/verb, which one?
(*interpolated function, interpolate, interpolation?*)
- constructions with two or more adjectives, add sub-expressions
differential graded category, graded category?
- expressions with prepositions, e.g. *sheaf of germs of analytic functions*
- proper nouns in adjective-like positions as *Shanin's method*
Lagrange interpolation

- Treat math concepts as black-boxes, as possible
- Use singular for concepts
- convention: Terms like theorem, corollary, conjecture, paper are not concepts
- convention: Mathematicians are not mathematical concepts
- If a long span is a concept, add important subspans e.g. *enriched accessible category*, *accessible category*
- avoid prepositions inside of concepts *sheaf of germs of analytic functions* we add instead *sheaf*, *germ*, *analytic function*

- Both TAC abstracts and nLab entries already processed with spaCy and available at Topos github <https://github.com/ToposInstitute/nlab-corpus>.
- Instead of using an ATE system we want to see how chatGPT can do in the task of extracting concepts
- Use sanitized sentences, i.e. sentences of medium length and with no \LaTeX , after LateXML (Miller).
- Out of 3.2K TAC sentences around 400 satisfy those constraints. Out of 15K entries in the nLab c. 2020, 175151 sentences, 55K satisfy the constraints.

- Wanted three experiments: 100 sentences Pilot, 436TAC, 55KnLab
- for Pilot100 we have three mathematicians and chatGPT is the fourth annotator
- For 436TAC only one annotator and chatGPT, plus adjudication
- For 55KnLab no human evaluation, so far

Given the following Context, extract the words that denote Math concepts.

Here are some examples:

{*in-context example*}

Now please solve the following problem.

Context: {*math_sentence*}

Concepts:

The initial prompt template, where *math_sentence* denotes the sentence from which we wish to extract math concepts.

- Pilot experiment: 100 sentences from TAC abstracts
- Examples of sentences:

We check these extra assumptions in several categories with pre-topologies.

We show that both approaches give equivalent bicategories.

In this paper, we use the language of operads to study open dynamical systems.

- Many 'minor mistakes' were made by the annotators
- NLI annotation tool repurposed

Math Concept Annotator

Upload Math Concept Adjudicate

436_sentences -

0

Start

-
-
-
-

Download

Results on experiment 1

Annotators Being Compared	Jaccard Score
annotator 1 and annotator 2	0.753
annotator 1 and annotator 3	0.794
annotator 2 and annotator 3	0.746
ChatGPT and annotator 1	0.485
ChatGPT and annotator 2	0.518
ChatGPT and annotator 3	0.505
ChatGPT and union of the humans	0.45
ChatGPT (after filtering) and union of the humans	0.5

Figure 3: Comparison of Jaccard similarities in experiment 1.

Results on experiment 2

Measurement	Jaccard Score
Between Human and Un-filtered ChatGPT	0.531
Between Human and Filtered ChatGPT	0.631

- problems of mathematical term extraction
- inter-annotator (dis)agreements
- guidelines to standardize the process
- an annotation tool to help humans with ATE
- best practices for prompts to ChatGPT
- discussing whether ChatGPT could be used as an annotator similar to a human expert

Preliminary Contributions

- Results over surveyable collections of data are not too good
- With judicious prompting, ChatGPT can find math terms, but misses some important ones
- Also it extracts terms that are not “mathematical terms” e.g. *conjecture*
- Also hallucinations (e.g Dress-Mueller Theorem)
- ChatGPT may be used very easily and gives acceptable preliminary results. Hence it WILL.
- Need to measure the quality of these LLMs results
- we have first numbers, hopefully we can build on these

What's a group?

Definition [\[edit \]](#)

The axioms for a group are short and natural... Yet somehow hidden behind these axioms is the [monster simple group](#), a huge and extraordinary mathematical object, which appears to rely on numerous bizarre coincidences to exist. The axioms for groups give no obvious hint that anything like this exists.

[Richard Borcherds](#), *Mathematicians: An Outer View of the Inner World*^[4]

A group is a non-empty [set](#) G together with a [binary operation](#) on G , here denoted " \cdot ", that combines any two [elements](#) a and b of G to form an element of G , denoted $a \cdot b$, such that the following three requirements, known as [group axioms](#), are satisfied:^{[5][6][7][a]}

But before we even get there, \mathbb{Z} .

1. Definition

Classically, a [group](#) is a [monoid](#) in which every element has an [inverse](#) (necessarily unique). When written with a view toward [group objects](#) (see Internalization below), one should rather say that a group is a monoid together with an inversion operation.

An [abelian group](#) is a group in which moreover the order in which two elements are multiplied is irrelevant.

2. Delooping

To some extent, a group "is" a [groupoid](#) with a single object, or more precisely a [pointed](#) groupoid with a single object.

What is a group?

According to Lean,

```
@[class]
structure group (G : Type u) :
  Type u
  (mul : G → G → G)
  (mul_assoc : ∀ (a b c : G), a * b * c = a * (b * c))
  (one : G)
  (one_mul : ∀ (a : G), 1 * a = a)
  (mul_one : ∀ (a : G), a * 1 = a)
  (npow : ℕ → G → G)
  (npow_zero' : (∀ (x : G), group.npow 0 x = 1) . "try_refl_tac")
  (npow_succ' : (∀ (n : ℕ) (x : G), group.npow n.succ x = x * group.npow n x) . "try_refl_tac")
  (inv : G → G)
  (div : G → G → G)
  (div_eq_mul_inv : (∀ (a b : G), a / b = a * b⁻¹) . "try_refl_tac")
  (zpow : ℤ → G → G)
  (zpow_zero' : (∀ (a : G), group.zpow 0 a = 1) . "try_refl_tac")
  (zpow_succ' :
    (∀ (n : ℕ) (a : G), group.zpow (int.of_nat n.succ) a = a * group.zpow (int.of_nat n) a) .
    "try_refl_tac"
  )
  (zpow_neg' :
    (∀ (n : ℕ) (a : G), group.zpow -(1 + n) a = (group.zpow +(n.succ) a)⁻¹) . "try_refl_tac")
  (mul_left_inv : ∀ (a : G), a⁻¹ * a = 1)

A group is a monoid with an operation  $^{-1}$  satisfying  $a^{-1} * a = 1$ .

There is also a division operation  $/$  such that  $a / b = a * b^{-1}$ , with a default so that  $a / b = a * b^{-1}$  holds by definition.
```

- Many resources, many audiences
 - How can people from different backgrounds make sense of them?
- NYT calls assistants "proof whiners" for not understanding definitions ("AI is Coming for Mathematics, Too")
 - How can we improve communication between mathematician and computer?

- Organize the various resources for undergraduate-level math
- Use this organization to:
 - make math more accessible to everyone: students, non-mathematician scientists, "laypeople," etc.
 - bridge the language gap between natural math and formal

MathGloss

This table contains Wikidata ID numbers and the corresponding term from each corpus. To see terms from the corpora that were not mapped to Wikidata, click on the links in the headers.

Wikidata ID	Chicago	France UG	nLab	Hosgood
Q14481419		0-1 law		
Q181296	abelian		abelian group	abelian group
Q318598	abelianization		abelianization	
Q20827138	absolute continuity of measure			
Q332504	absolute continuity		absolutely continuous measure	
Q332465		absolute convergence	absolute convergence	absolute convergence
Q120812	absolute value		absolute value (mathematics)	absolute value (mathematics)

Our table, combining terms from four (more to come!) different corpora and linking them to Wikidata. So far, 906 terms in total.

Example

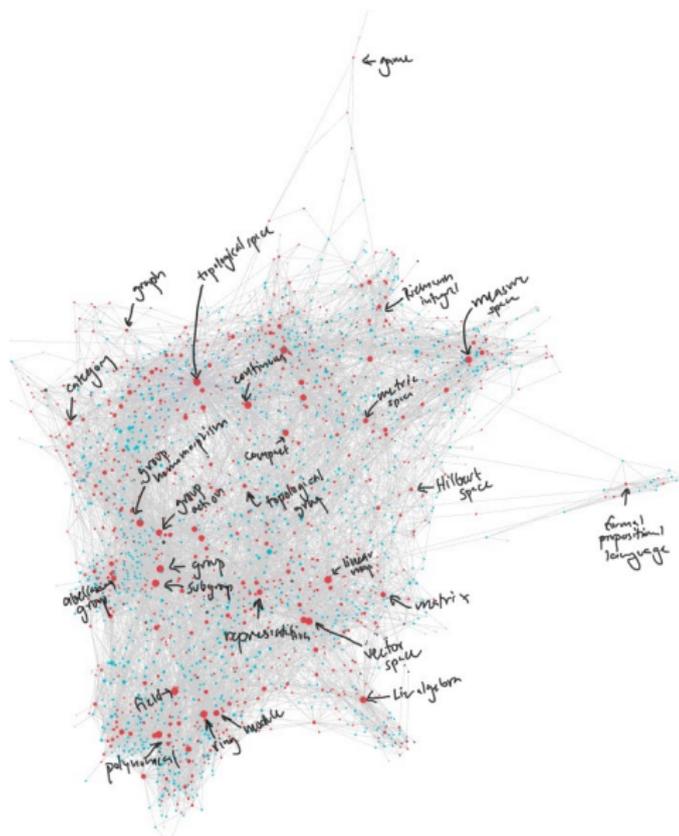
bers and the corresponding term from each corp
e not mapped to Wikidata, click on the links in the

Chicago	France UG	nLal
	0-1 law	
an		abelian gro
anization		abelianizat

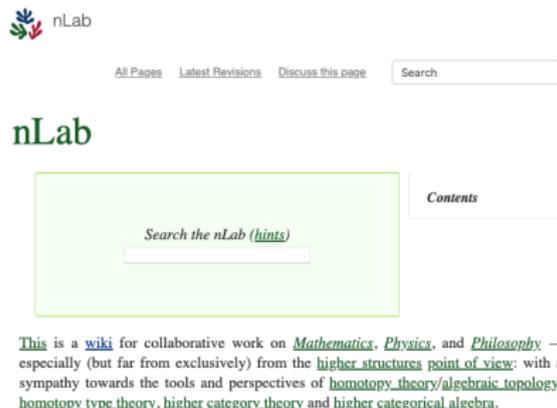
```
mathlib3 probability.independence.zero_one source
theorem probability_theory.
  measure_zero_or_one_of_measurable_set_limsup_at_top
  (Ω : Type u_1) (ι : Type u_2) (m0 : measurable_space Ω)
  (μ : measure_theory.measure Ω)
  [measure_theory.is_probability_measure μ]
  (s : ι → measurable_space Ω) [semilattice_sup ι]
  [no_max_order ι] [nonempty ι] (h_le : ∀ (n : ι), s n ≤ m0)
  (h_indep : probability_theory.Indep s μ) (t : set Ω)
  (ht_tail : measurable_set t) :
  ↑μ t = 0 ∨ ↑μ t = 1
Kolmogorov's 0-1 law : any event in the tail σ-algebra of an independent sequence of
sub-σ-algebras has probability 0 or 1. The tail σ-algebra  $\limsup s$  at_top is the
same as  $\bigcap_n \bigcup_{i \geq n} s_i$ .
```

Chicago corpus

- (almost) all notes from my math major @UChicago, atomized and linked
- Theorems are not yet included in MathGloss
- 758 definitions total, 500 mapped manually to Wikidata



- Page *titles* from the nLab, a higher-math wiki, minus people
- 18k+ pages, 5377 mapped to Wikidata with Wikimapper
- We restrict to those terms which are in other corpora



The screenshot shows the nLab website interface. At the top left is the nLab logo, a stylized tree with blue, green, and red leaves. To its right is the text "nLab". Below the logo are three links: "All Pages", "Latest Revisions", and "Discuss this page". To the right of these links is a search input field with the placeholder text "Search".

Below the navigation links is the heading "nLab" in a large, bold, green font. Underneath this heading is a light green rectangular box containing the text "Search the nLab ([hints](#))" and a search input field.

To the right of the search box is a "Contents" link. Below the search box is a paragraph of text: "This is a [wiki](#) for collaborative work on [Mathematics](#), [Physics](#), and [Philosophy](#) — especially (but far from exclusively) from the [higher structures point of view](#); with a sympathy towards the tools and perspectives of [homotopy theory/algebraic topology](#), [homotopy type theory](#), [higher category theory](#) and [higher categorical algebra](#)."

- Similar to MathGloss but with cross-language rather than cross-resource linking
- 305 terms in total, some languages don't have their word included for every concept
- Terms are already mapped to Wikidata

Reference	EN	FR	JA
0318737	abelian category	catégorie abélienne (<i>f</i>)	アーベル圏
0181296	abelian group	groupe abélien (<i>m</i>)	アーベル群
0515874	abscissa	abscisse (<i>f</i>)	
0120812	absolute value	valeur absolue (<i>m</i>)	
091134251	absolutely convergent series	série absolument convergente (<i>f</i>)	
0844451	acnode	point isolé (<i>m</i>)	
03250296	acute angle	angle aigu (<i>m</i>)	
032043	addition	addition (<i>f</i>)	加法
04681343	additive category	catégorie additive (<i>f</i>)	加法圏
0320346	adherence	adhérence (<i>f</i>)	

Where else?

- Other online databases
- Use NLP to get terms from other types of resources like textbooks:
 - TAC stands for "Theory and Applications of Categories," a journal in category theory
 - We took the 755 abstracts from papers and processed the text using spaCy (an NLP library for Python) and extracted "concepts:"
 - nouns, compounds, adjective-noun phrases
 - selected for frequency of appearance
 - Not included in this iteration of MathGloss

Other interactive theorem provers could produce resources like the UG \Rightarrow Lean4.

What's Next?

- Three parallel projects are progressing.
- Each has new milestones
- More corpora would be nice
- Golden standards are part of the plan
- We need a KG plan– as off-the-shelf as possible
- Also advancing the theory to NLI and (linguistic) causality

Funding would help!

Some References



Collard et al, *Extracting Mathematical Concepts from Text*. In W-NUT 2022, arxiv 2208.13830.



Collard et al, *Parmesan: mathematical concept extraction for education* arxiv 2307.06699.



de Paiva et al, *Extracting Mathematical Concepts with Large Language Models* arXiv:2309.00642



Horowitz, de Paiva, *MathGloss: Linked Undergraduate Math Concepts*, EuroProofNet Workshop 2023.



spaCy Industrial-Strength Natural Language Processing,
<https://spacy.io/>



Universal Dependencies, <https://universaldependencies.org/>



LateXML, <https://math.nist.gov/~BMiller/LaTeXML/>



Math Concept Annotator,
https://gaoq111.github.io/math_concept_annotation/

Model	Entities	Relations
PL-Marker	0.699	0.532
SpERT.PL	0.7053	0.5125
SpERT	0.703	0.5084
Cross-Sentence	0.681	0.501
DyGIE++	0.675	0.484

Table 1: F_1 scores for several models on the SciERC benchmark

Model	Precision	Recall	F_1
DyGIE++	0.22	0.35	0.27
MWE	0.12	0.78	0.20
PL-Marker	0.23	0.38	0.28
SpERT.PL	0.14	0.77	0.23
Textrank	0.15	0.55	0.23

Table 2: Terminology extraction algorithms evaluated against TAC author keywords

Model	Precision	Recall	F_1
DyGIE++	0.12	0.27	0.16
MWE	0.08	0.75	0.14
PL-Marker	0.11	0.27	0.16
SpERT.PL	0.08	0.63	0.14
Textrank	0.09	0.46	0.14

Table 3: Terminology extraction algorithms evaluated against nLab titles appearing in TAC