MathGloss: Linked Undergraduate Math Concepts

Lucy Horowitz Valeria de Paiva

EuroProofNet Workshop 2023

September 7, 2023

What is a group?

What is a group?

According to Wikipedia,

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, $\mathbb{Z}$.

What is a group?

According to Wikipedia,

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, $\mathbb{Z}$.

According to the nLab,

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,

source

```
@[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.

The Situation

The Situation

- Many resources, many audiences

The Situation

- Many resources, many audiences
  - How can people from different backgrounds make sense of them?

The Situation

- 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")

The Situation

- 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?

Goals

- Organize the various resources for undergraduate-level math

Goals

- Organize the various resources for undergraduate-level math
- Use this organization to:

Goals

- Organize the various resources for undergraduate-level math
- Use this organization to:
    - make math more accessible to everyone: students, non-mathematician scientists, "laypeople," etc.

Goals

- 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

Solution: MathGloss

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 corpu
e not mapped to Wikidata, click on the links in the

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

Example

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

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

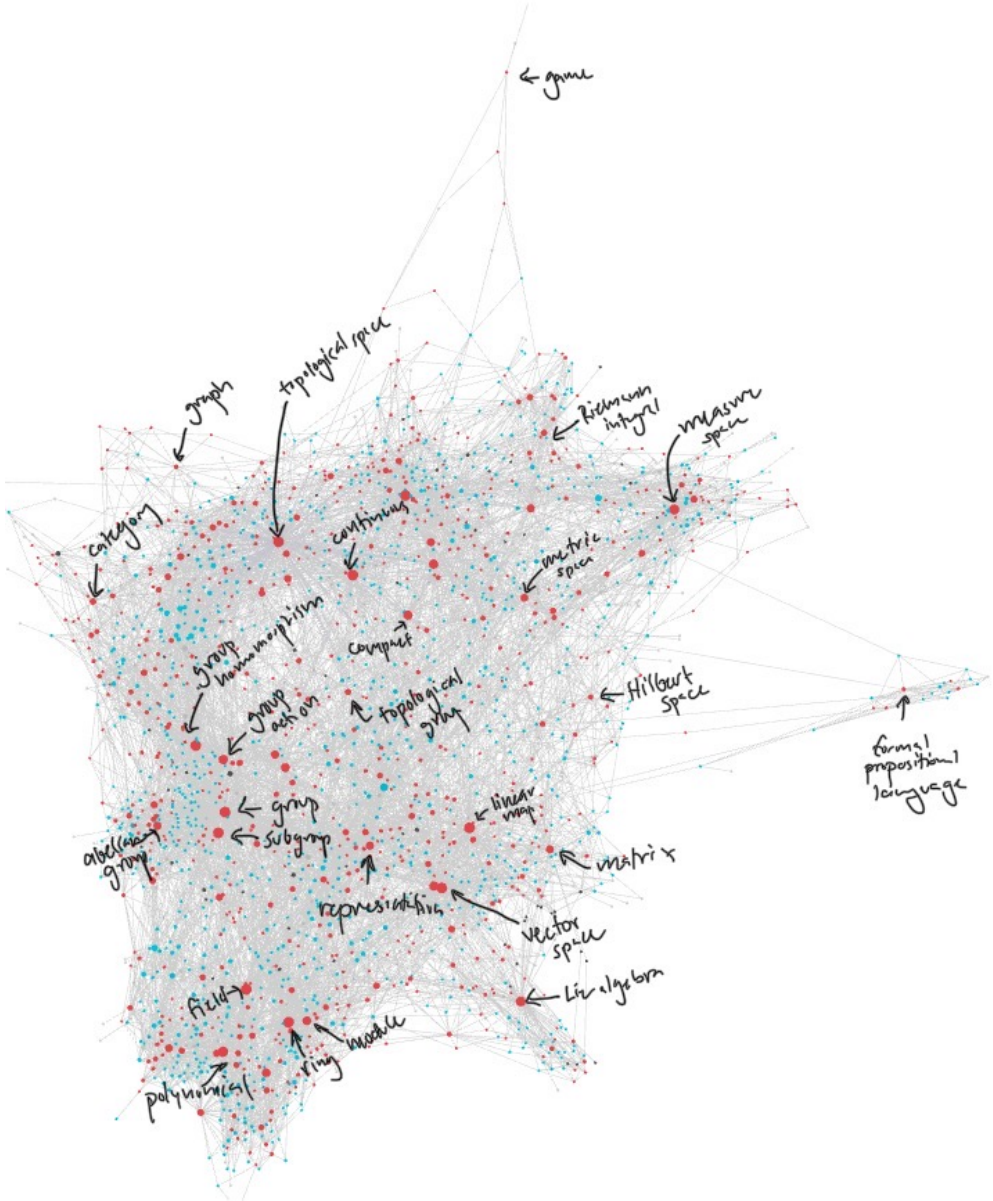
mathlib3 probability.independence.zero_one

```
theorem probability_theory.                                    source
    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 `⋂ n, ⋃ i ≥ 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

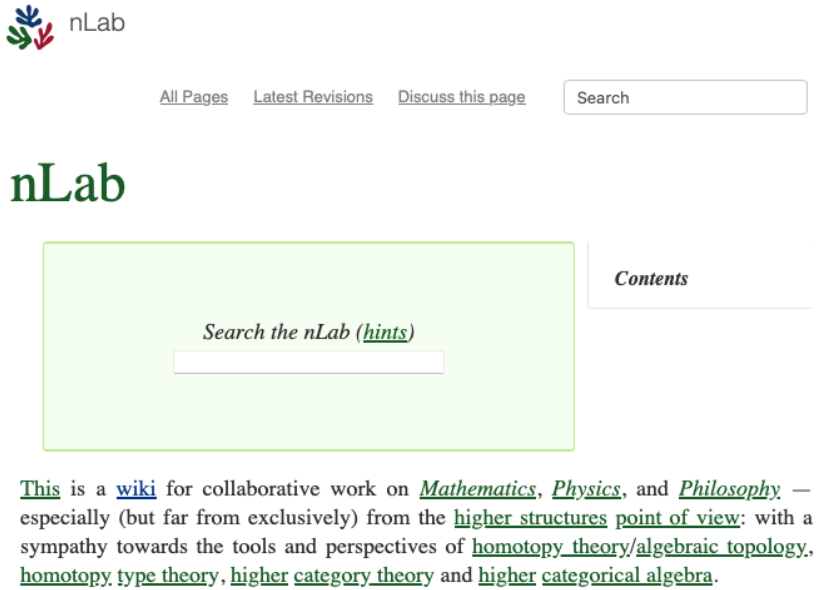


France UG corpus

- Concepts the French government expects its undergrads to know, organized by subject
- Terms are translated into English
- Where possible, a pointer is given to the corresponding entry in Lean
- 543 terms total, 369 mapped to Wikidata with `wikimapper`

nLab corpus

- 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



Tim Hosgood's dictionary

- 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
Q318737	abelian category	catégorie abélienne *(f)*	アーベル圏
Q181296	abelian group	groupe abélien *(m)*	アーベル群
Q515874	abscissa	abscisse *(f)*	
Q120812	absolute value	valeur absolue *(m)*	
Q91134251	absolutely convergent series	série absolument convergente *(f)*	
Q844451	acnode	point isolé *(m)*	
Q3250296	acute angle	angle aigu *(m)*	
Q32043	addition	addition *(f)*	加法
Q4681343	additive category	catégorie additive *(f)*	加法圏
Q320346	adherence	adhérence *(f)*	

Jan-Christoph Klie's `wikimapper`

Jan-Christoph Klie's `wikimapper`

- A simple Python package that takes a page title and spits out a Wikidata ID

Jan-Christoph Klie's `wikimapper`

- A simple Python package that takes a page title and spits out a Wikidata ID
- But there is the common problem of math words having the same spelling as regular English words

Jan-Christoph Klie's `wikimapper`

- A simple Python package that takes a page title and spits out a Wikidata ID
- But there is the common problem of math words having the same spelling as regular English words
    - e.g. group, field, ring...

Jan-Christoph Klie's `wikimapper`

- A simple Python package that takes a page title and spits out a Wikidata ID
- But there is the common problem of math words having the same spelling as regular English words
    - e.g. group, field, ring...
- `wikimapper` often maps to "disambiguation" pages, when Wikipedia itself knows that there are multiple senses of the same word

Jan-Christoph Klie's `wikimapper`

- A simple Python package that takes a page title and spits out a Wikidata ID
- But there is the common problem of math words having the same spelling as regular English words
    - e.g. group, field, ring...
- `wikimapper` often maps to "disambiguation" pages, when Wikipedia itself knows that there are multiple senses of the same word
- We circumvent this by using Wikipedia's system: add (mathematics) or (topology) etc. to terms, and *then* map

Where else?

Where else?

- Other online databases

Where else?

- Other online databases
- Use NLP to get terms from other types of resources like textbooks:

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

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

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

Future work

Future work

- Pick up spaCy again, along with the problem of "definition extraction" from open-source textbooks

Future work

- Pick up spaCy again, along with the problem of "definition extraction" from open-source textbooks
- Expand to more online resources with even more viewpoints

Future work

- Pick up spaCy again, along with the problem of "definition extraction" from open-source textbooks
- Expand to more online resources with even more viewpoints
- Make the site a bit easier to navigate!

Future work

- Pick up spaCy again, along with the problem of "definition extraction" from open-source textbooks
- Expand to more online resources with even more viewpoints
- Make the site a bit easier to navigate!
- Make the mappings easy to update

Some References

(see https://mathgloss.github.io/MathGloss/)

- The nLab: ncatlab.org
- Undergraduate mathematics in mathlib: https://leanprover-community.github.io/undergrad.html
- Tim Hosgood's Dictionary: https://thosgood.com/maths-dictionary/
- `wikimapper`: https://github.com/jcklie/wikimapper
- Roberts, Siobhan. "AI Is Coming for Mathematics, Too": https://www.nytimes.com/2023/07/02/science/ai-mathematics-machine-learning.html