

Exploratory analysis of topic co-occurrence in the web counselling service

.....

Marjan Cugmas

Faculty of Social Sciences, University of Ljubljana

Domen Kralj

National Institute of Public Health

Applied Statistics 2024



This is me

The online counselling service was established in April 2001.

The service works as a moderated Web forum where users (adolescent) post their questions, and qualified experts post their answers.

The program is recognized as a best practice by the OECD.

100 VOLUNTEER EXPERTS

Psychologists, medical specialists and experts in other specialties.

37k ANSWERS IN THE PAST 23 YEARS

In their answers, the experts encourage positive changes, motivate, direct, listen, relieve, are optimists and informers.

200k VISITS IN 2023

According to Google Analytics, most visitors are "new", which suggests that they visit the counselling service especially when they need advice.



Communication flow

This is the moderated online health community.



Users

Write and submit a question on the web.



Editorial Office

Make the question public (or not).
Select a counselor and send them the question.
Post the answer when received.

Since 2012, manually assign gender, age, and topics based on the codebook.



Counsellors

Write an answer and send it back to the Editorial Office.

The codebook: topics and subtopics

Multiple subtopics can be assigned to a single question.

Body	Physical maturation
	Congenital malformations
	Nutrition
	Exercising, shaping the body
	Body weight
	Adverse body self-image
	Body decoration
	Other (body)
Physical health	Cold, aches and pains, and minor illnesses
	Allergies
	Skin
	Eyes
	Oral health
	Hair, fingernails
	Serious illnesses
	Injuries
	Use of medicines
	Alternative medicine
	Sleep
	Other (physical health)

Sexuality	Anatomy of genitals
	Sexual needs, drive
	Kissing
	Masturbation
	Petting
	Orgasm
	Sexual relationships
	Deciding to have sex
	Online sex behaviour
	Sex games, toys, fetishes
	Sexual orientation and identity
	Other (sex/sexuality)
Sexual health	Contraception
	Sexually transmitted diseases
	Reproduction
	Termination of pregnancy
	Teen pregnancy
	Women's sexual health
	Gynaecological examination
	Men's sexual health
	Safe sex
	Fear of pregnancy
	Pregnancy test
	Protection planning
	Contraception in men
	Contraception in women
	Post-coital contraception
	Genitals
	Other (sexual health)

Relationships	Love
	Relationships in family
	Relationships with peers
	Deviant behaviour
	Other (relationships)
	Relationships at home
	Parental questions
	Friendships
	Friendships with benefits
	Relationships with peers
	Long-distance/online relationships
	Peer violence
	Domestic violence
	Sexual abuse
	Delinquency, crime
Mental health	Communication problems
	Stress
	Emotions
	Self-image
	Eating disorders
	Self-harm
	Mourning, losses
	Finding a consultant
	Anxiety
	Depression
	Suicidality
	Psychiatric problems
	Other (mental health)

Drugs, Addiction	Alcohol
	Nicotine
	Marijuana
	Party drugs
	Opiates, heroin, cocaine
	Inhalation drugs
	Misuse of prescribed medicines
	Misuse of laxatives
	Caffeine and energy drinks/food
	Relationship addiction
	Gambling addiction
	Computer addiction
School-related problems	Seeking help for addiction
	Other (drugs, addiction)
	Career orientation
	Learning difficulties
	Problems with teacher(s)
	Leisure time
Attitude towards society	Exclusion from school, sanctions
	Other (school-related problems)
	Mass culture, media
	Consumerism
	Poverty
	Standing up (advocacy) for something
	Worldview

Topics by years

The topics for the period 2001-2011 were determined using machine learning algorithms.

Number of questions

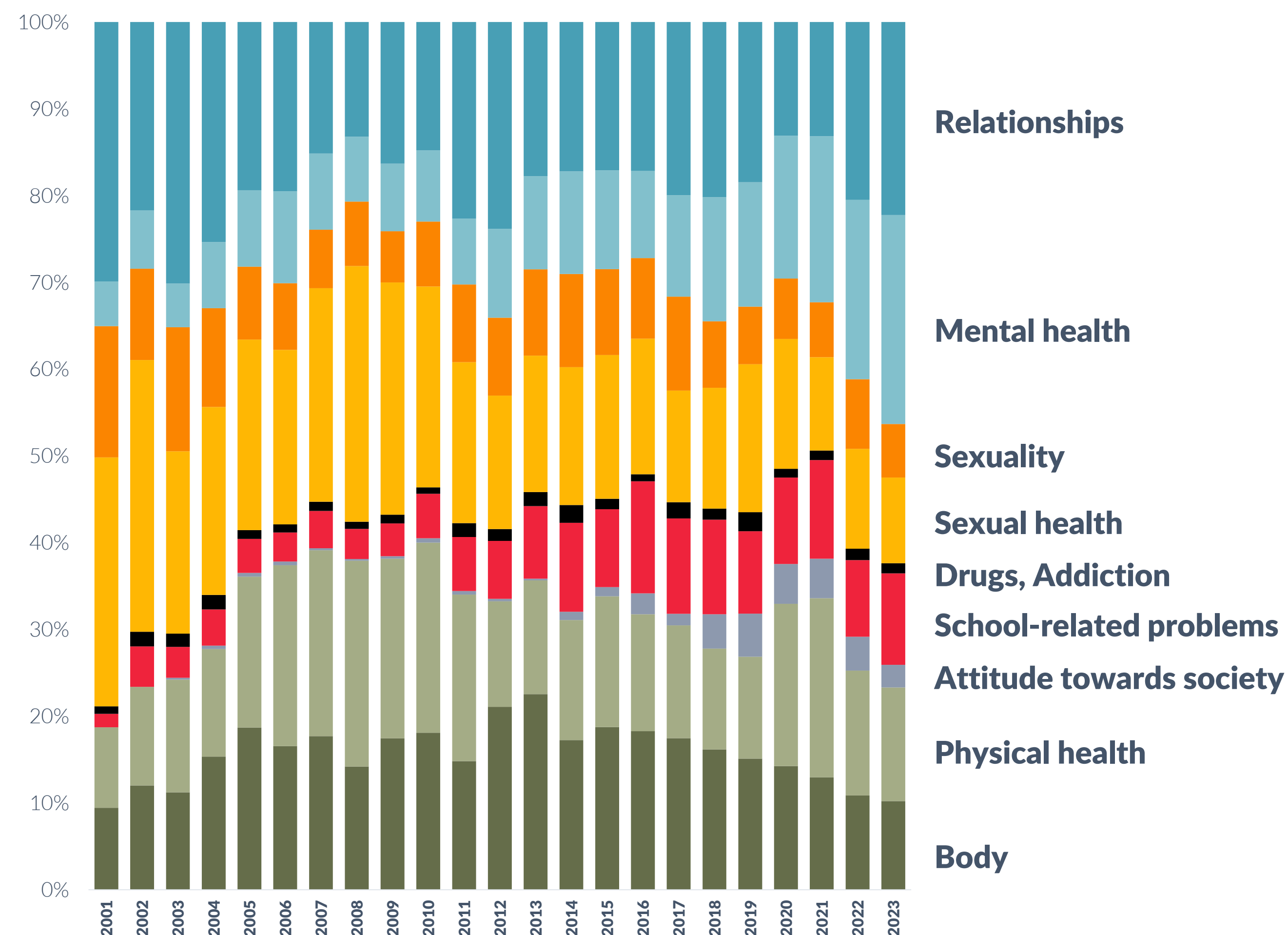
The number of questions is not subjected to the seasonal trends, but rather to the promotion activities.

- 2001-2013: 25,467 questions;
- 2014-2023: 12,039 questions.

The questions for the 2014-2023 period will be analyzed.

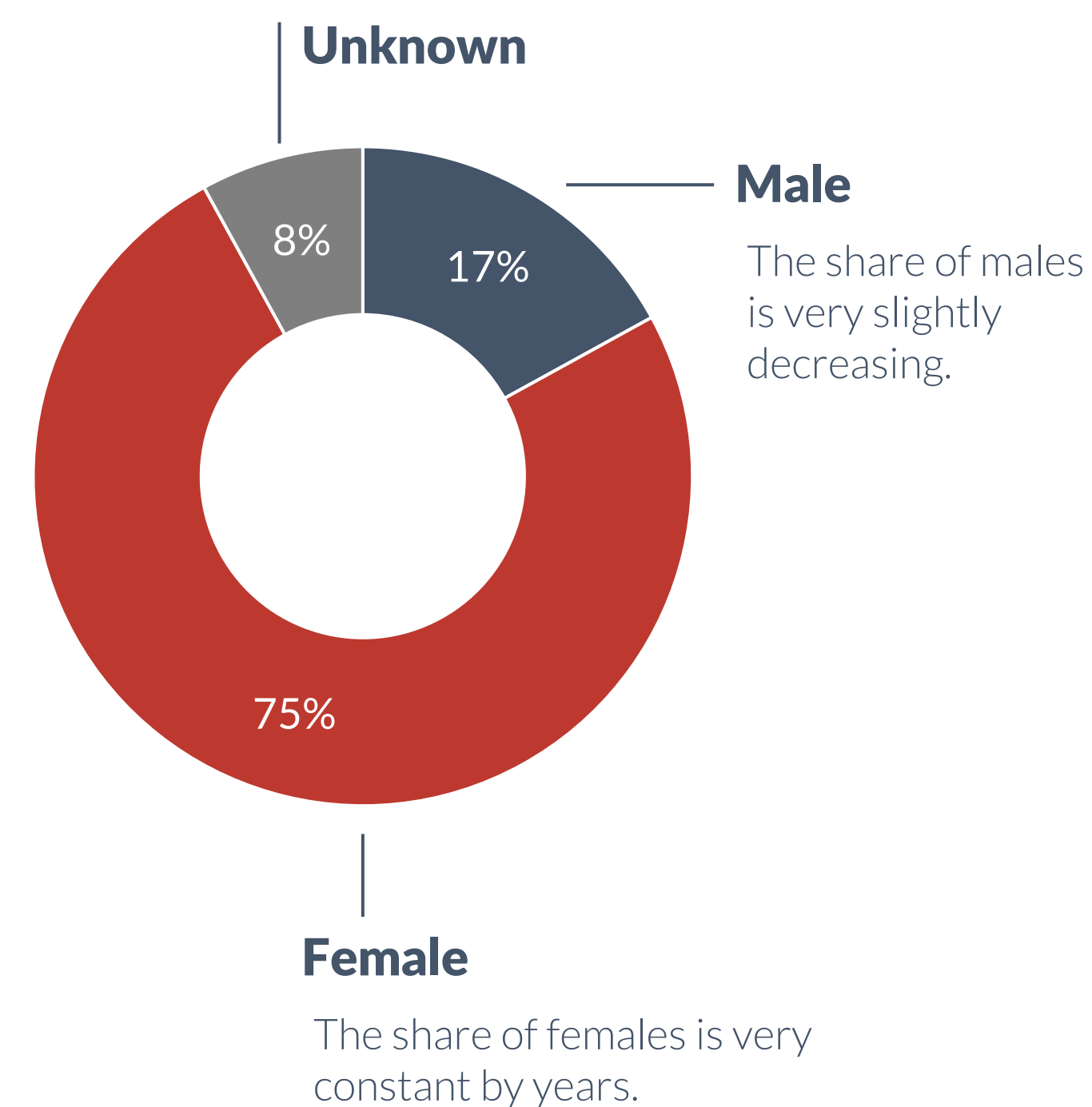
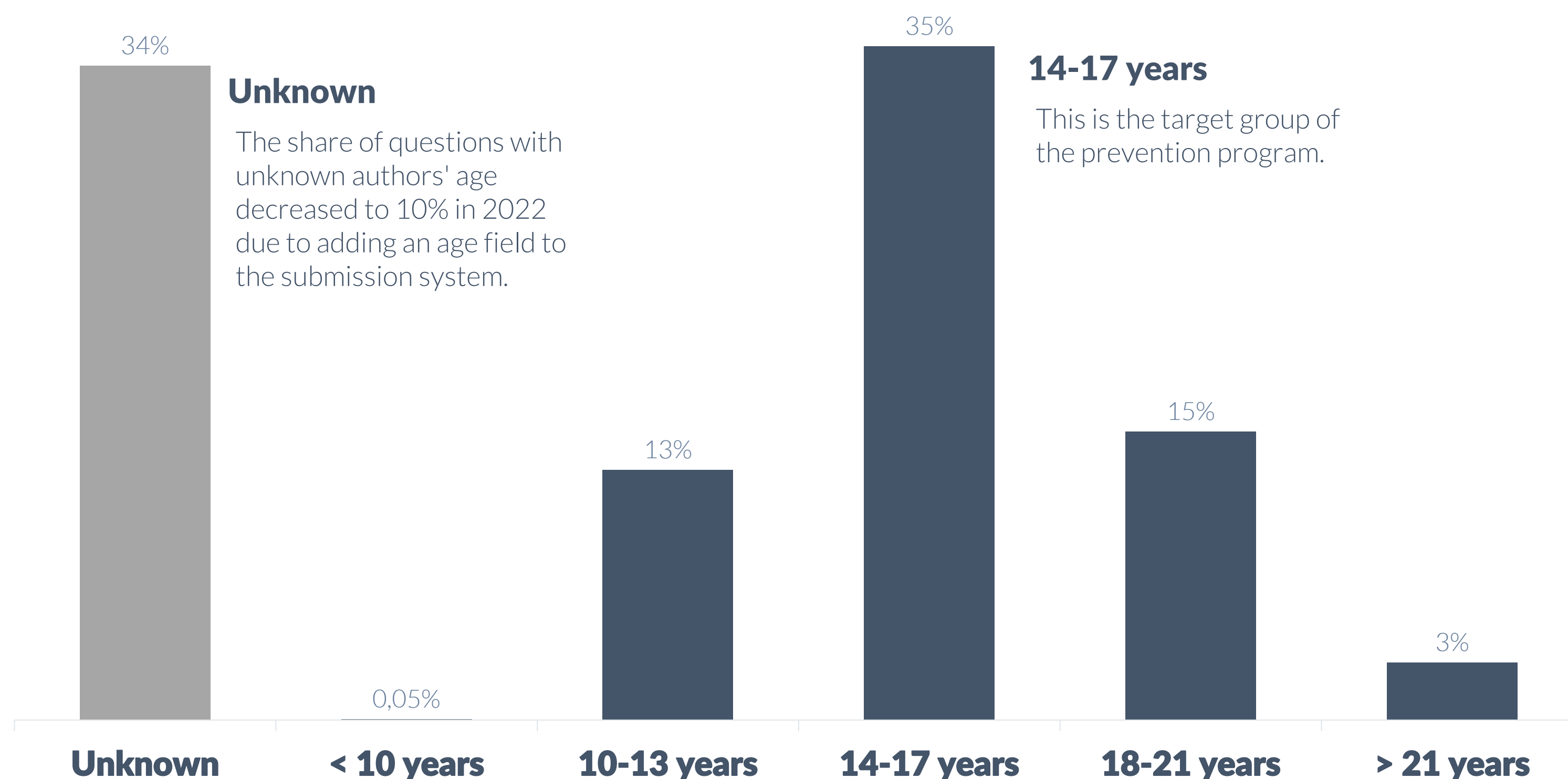
Topics

Questions related to relationships, mental health, and school issues are increasing, while those related to sexuality and sexual health are decreasing.



Age and gender structure

This refers to the characteristics of questions, which might differ from the user's demographics.



From maturation to reproduction

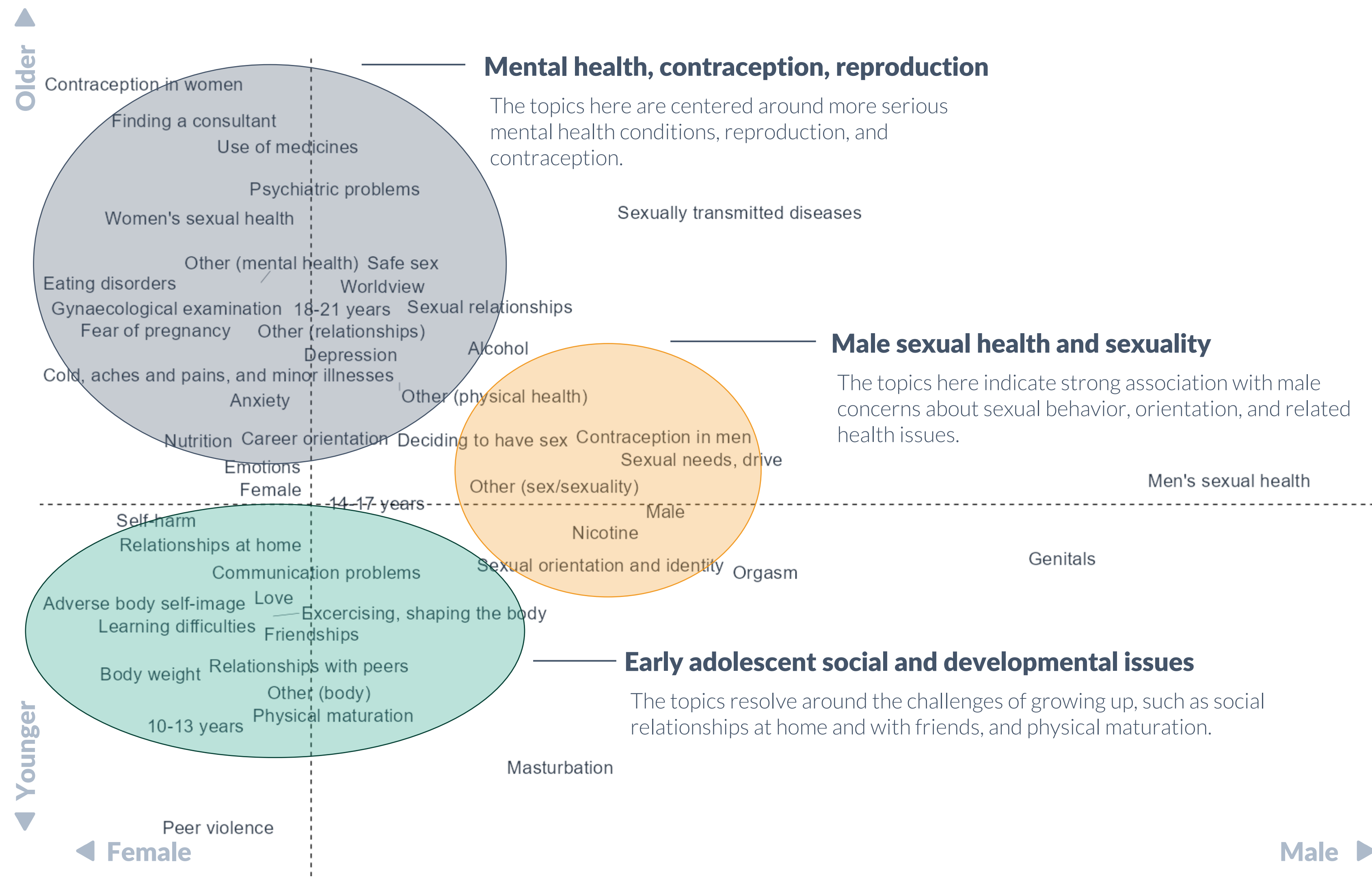
Exploring the interplay between topics, gender, and age groups.

MCA

The Multiple Correspondence Analysis with row-principal normalization was used to visualize associations between topics, gender, and age in a two-dimensional space. The most important 50 topics are visualized.

What we see?

Males are more focused on sexual health and identity. Younger females are more concerned with social relationships and body image, while older females shift their focus towards concerns related to mental health and sexual health and reproduction.



Communication – emotions – love

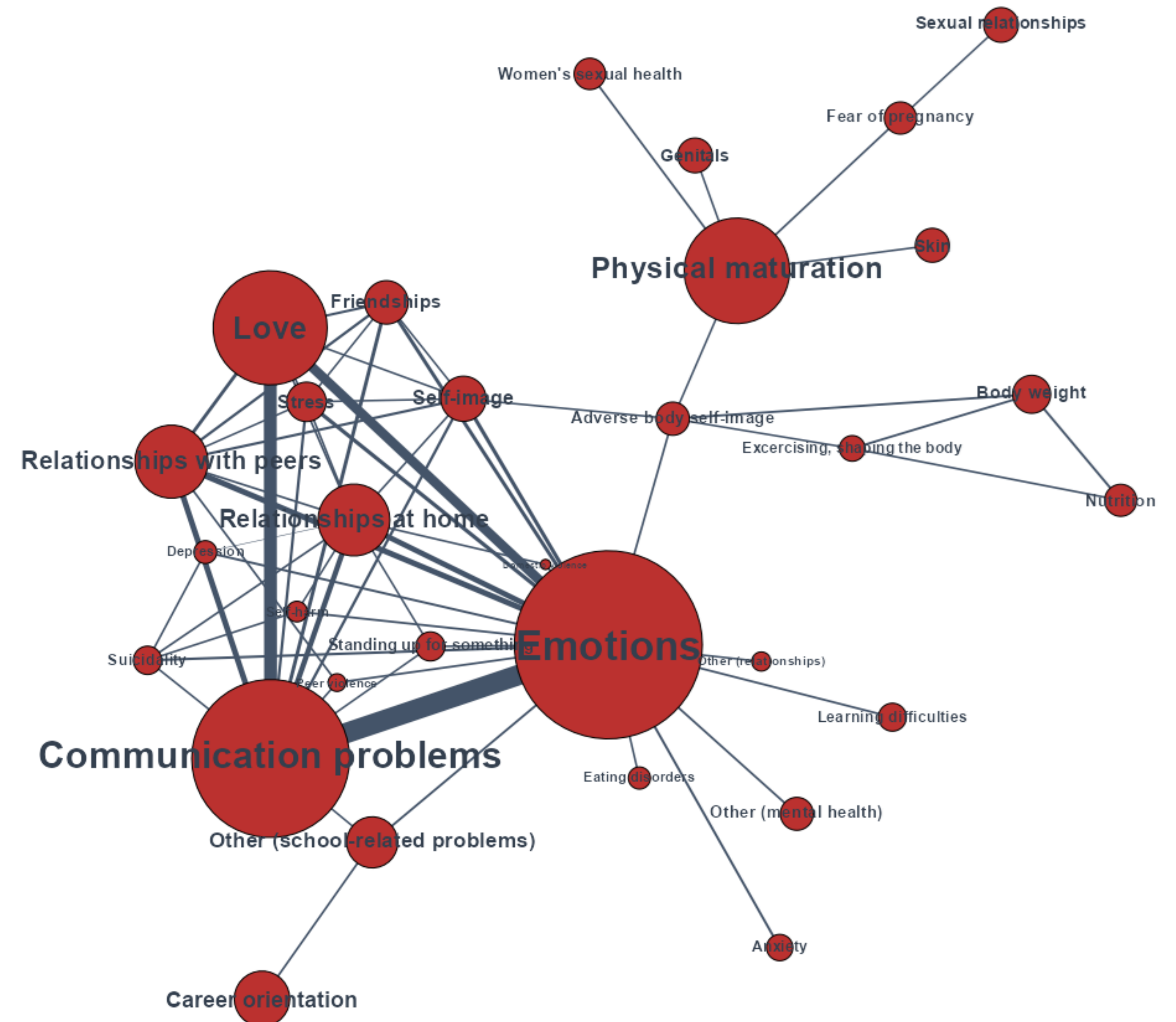
Nodes are topics and the link between them are co-occurrences.

Simple cut

Node sizes correspond to the number of topic occurrences, and edge widths are proportional to the number of co-occurrences. Only edges with more than 120 co-occurrences are preserved. This reflects the general impression of the editorial office on the topics in the web counseling service. Isolates are not shown (e.g., contraception in women, minor illnesses and other physical health).

What we see?

The communication – emotions – love triangle can be identified with some other typical issues related to adolescence, such as physical maturation, relationships with peers and relationships at home.



Topics overlap

Jaccard normalization reflects the topic overlap.

Aim

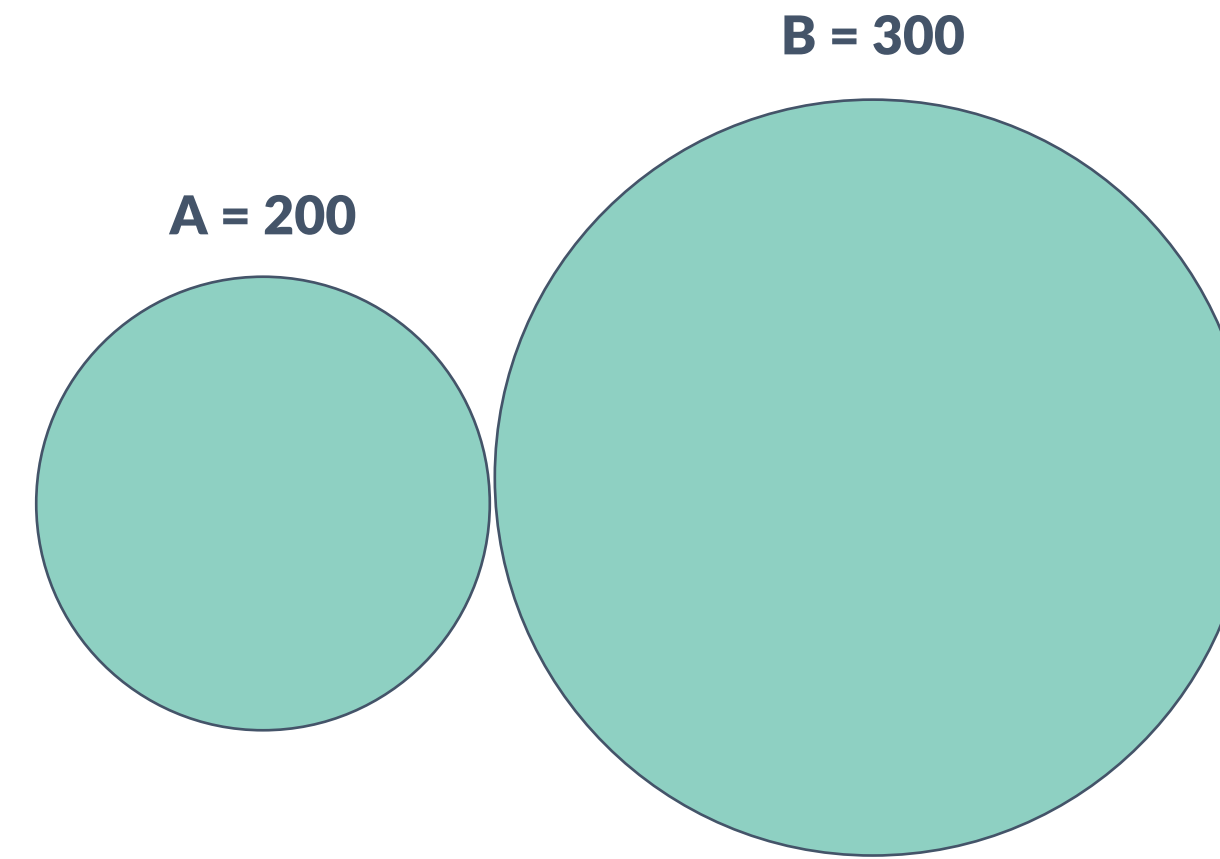
To reveal the more subtle topic co-occurrences by reducing the impact of the most common topics on the amount of topic co-occurrences.

Jaccard normalization

Given an adjacency network matrix of topic co-occurrences X , the Jaccard normalization for a given pair of topics (i and j) is defined as the ratio of the number of their co-occurrences to the number of co-occurrences with other topics, i.e.,

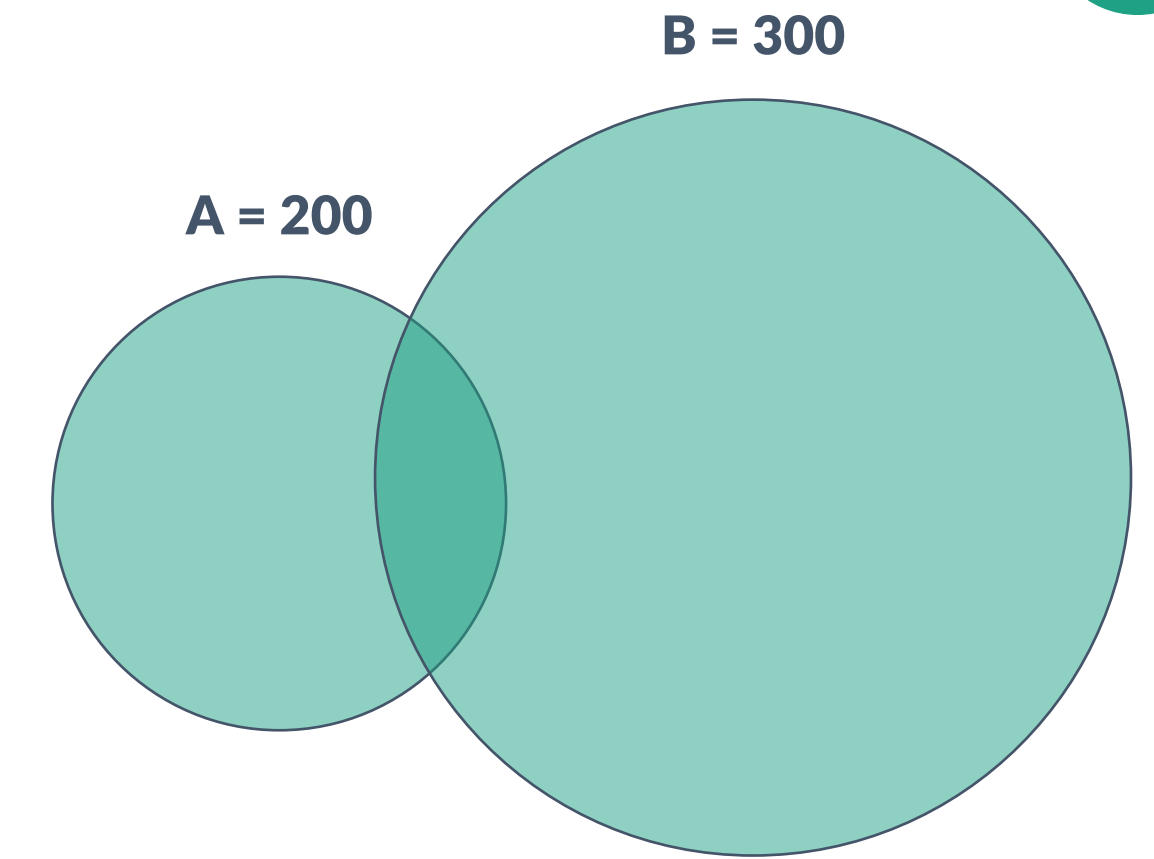
$$J = \frac{x_{ij}}{x_{i.} + x_{j.} - x_{ij}} \in [0, 1]$$

Note: this measure is still sensitive to large differences in topic appearances.



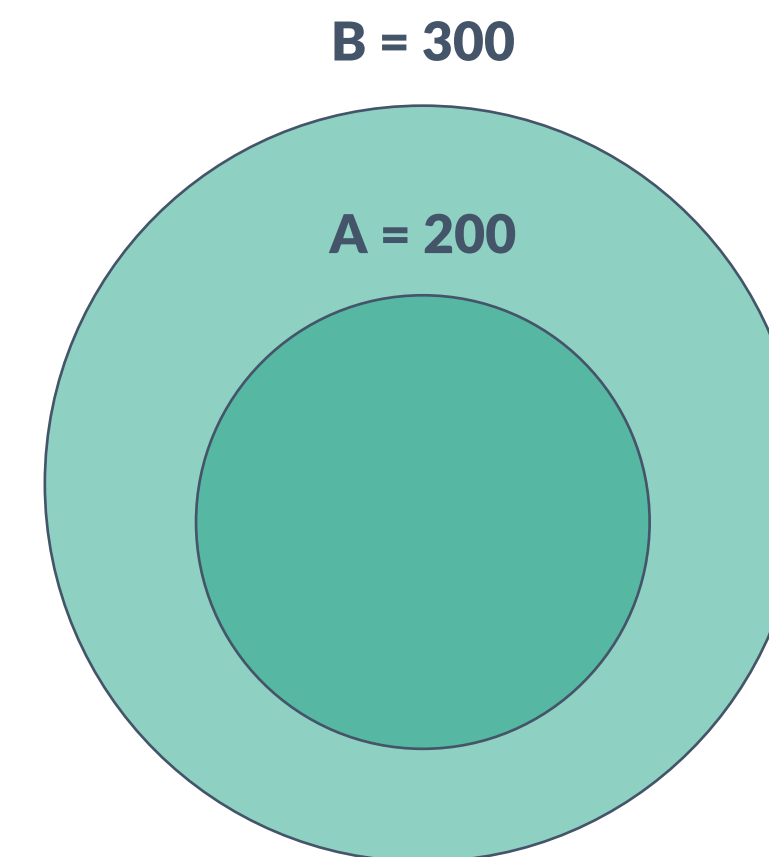
No overlap, $A \cap B = 0$, $J = 0$

The topics are completely separated, they never appear in the same question.



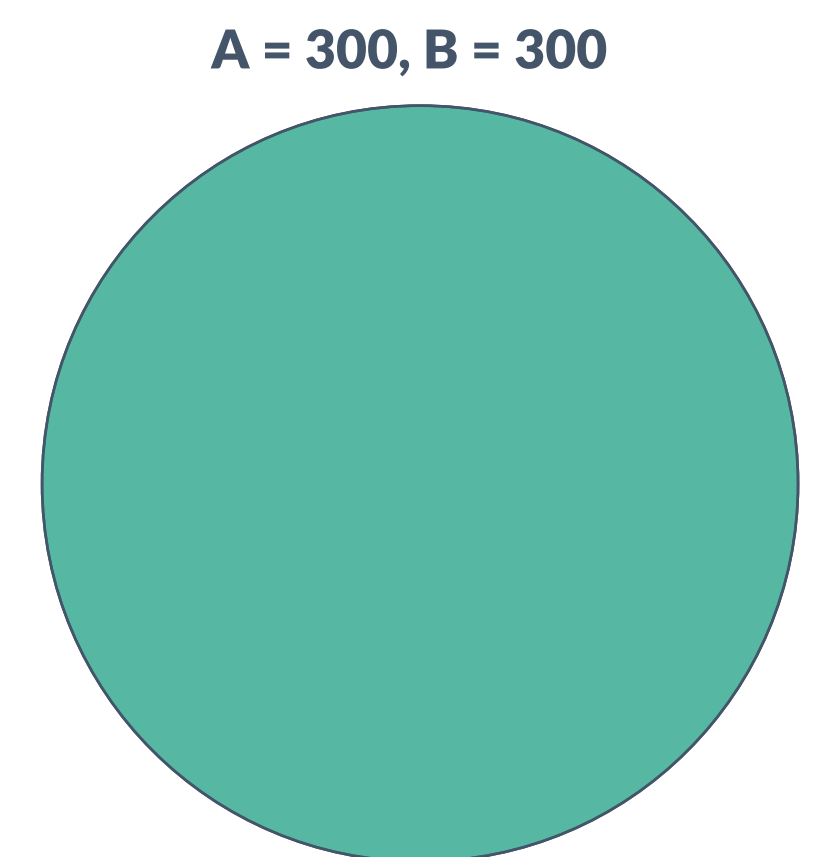
Overlap, $A \cap B = 100$, $J = 0.2$

Sometimes the topics co-appear in the same question.



Embedded, $A \cap B = 200$, $J = 0.7$

The topic A appears only with topic B, but since topic B also appear independently, the topic overlap is not complete.



Full overlap, $A \cap B = 300$, $J = 1.0$

The topics appear in the same questions, they never appear independently.

Five clusters were identified.

Sexual Health and Sexual Relationships

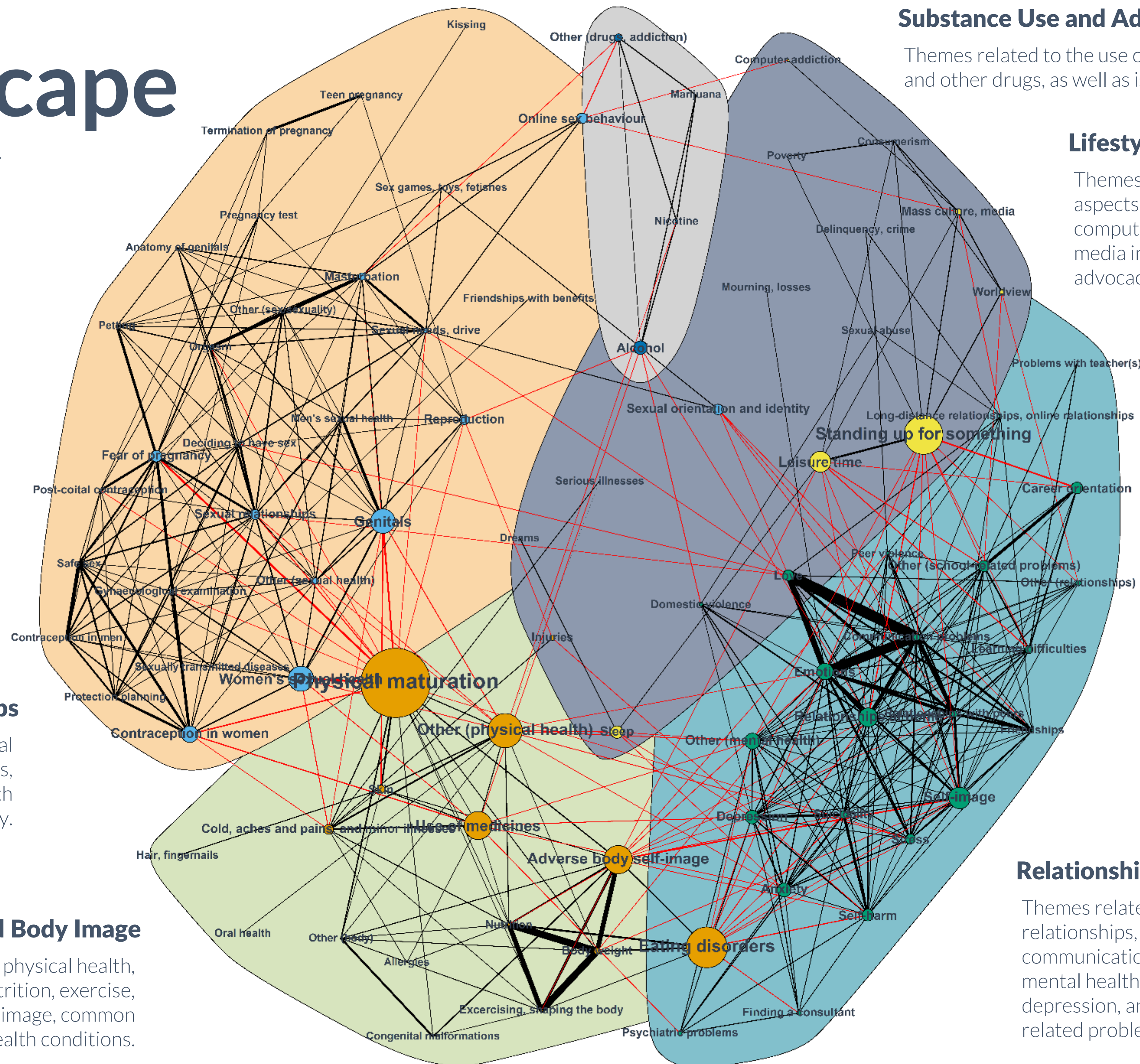
Physical Health and Body Image

Substance Use and Addiction

Lifestyle and Social Issues

Relationships and Mental Health

Themes related to love, family and peer relationships, domestic and peer violence, communication issues, emotional well-being, mental health challenges such as anxiety, depression, and suicidality, as well as school-related problems and career orientation.



Comparison with the original codebook

A high level of similarity is found, with substantively similar original umbrella topics mostly being merged.

The table below visualizes the comparison between the umbrella topics from the original codebook (rows) and the clusters identified based on a normalized network (columns). The number of common topics is shown in the cells of the table.



A high level of similarity is found with some minor differences. For example:

- Topic Friends with benefits might be closer to the sex-related issues than relationships.
- Topic Sleep might not be purely “physical category”.

	Physical Health and Body Image	Sexual Health and Sexual Relationships	Relationships and Mental Health	Substance Use and Addiction	Lifestyle and Social Issues
Body	7				
Physical health	9				1 (Sleep)
Sexual health		16			
Sexuality		12			
Mental health	1 (Eating disorders)		12		1 (Dreams)
Relationships		1 (Friends with benefits)	10		
School-related problems			4		1 (Leisure time)
Drugs, addiction				4	1 (Computer addiction)
Attitude toward society					5

Five clusters were identified.

Sexual Health and Sexual Relationships

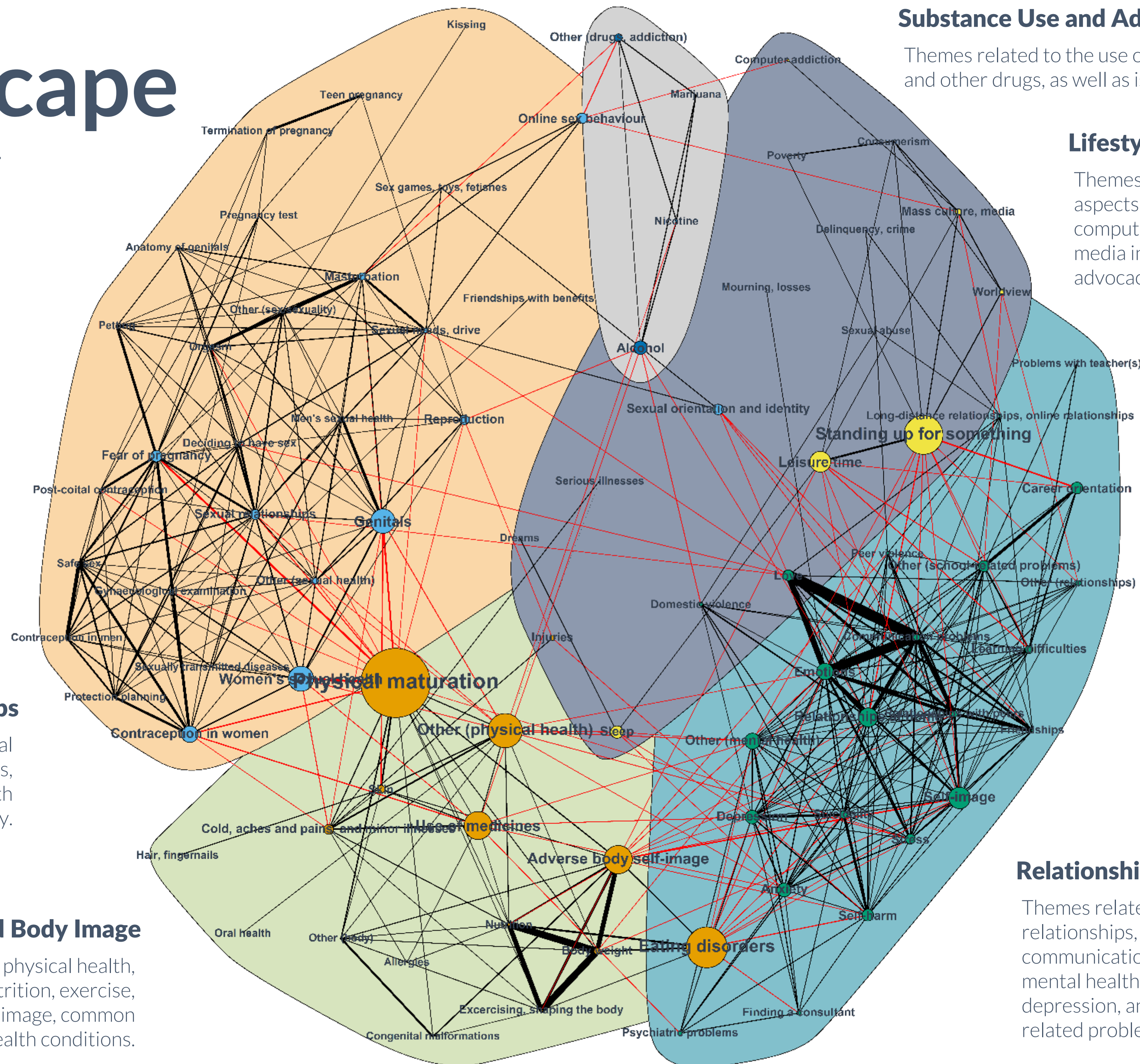
Physical Health and Body Image

Substance Use and Addiction

Lifestyle and Social Issues

Relationships and Mental Health

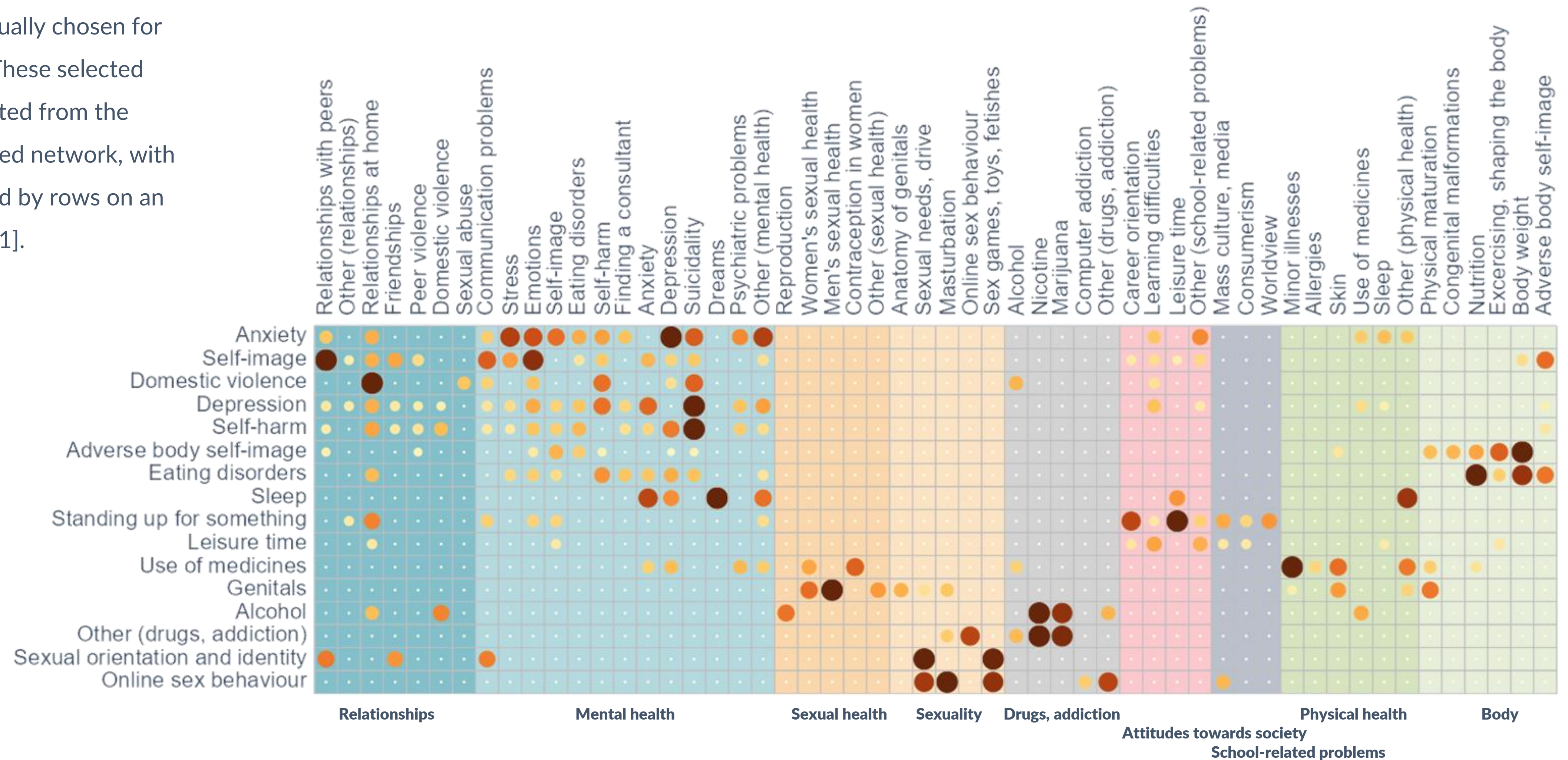
Themes related to love, family and peer relationships, domestic and peer violence, communication issues, emotional well-being, mental health challenges such as anxiety, depression, and suicidality, as well as school-related problems and career orientation.



Bridging topics

The topics that do not fit well into a single category.

Based on bridge strength, several topics were manually chosen for future analysis. These selected topics are extracted from the Jaccard normalized network, with values normalized by rows on an interval from [0, 1].



Insights

Three takeaway messages.



01

SURGE IN MENTAL HEALTH AND SOCIAL CONCERNS

The growing share of these questions may signal a rising demand for support in these areas and/or that other topics like sexuality are now more readily addressed through other channels.

02

HOW DO I TELL HIM THAT I LIKE HIM?

Questions related to growing up, such as physical maturation and relationships, remain the most prevalent. However, the specific topics are changing by age.

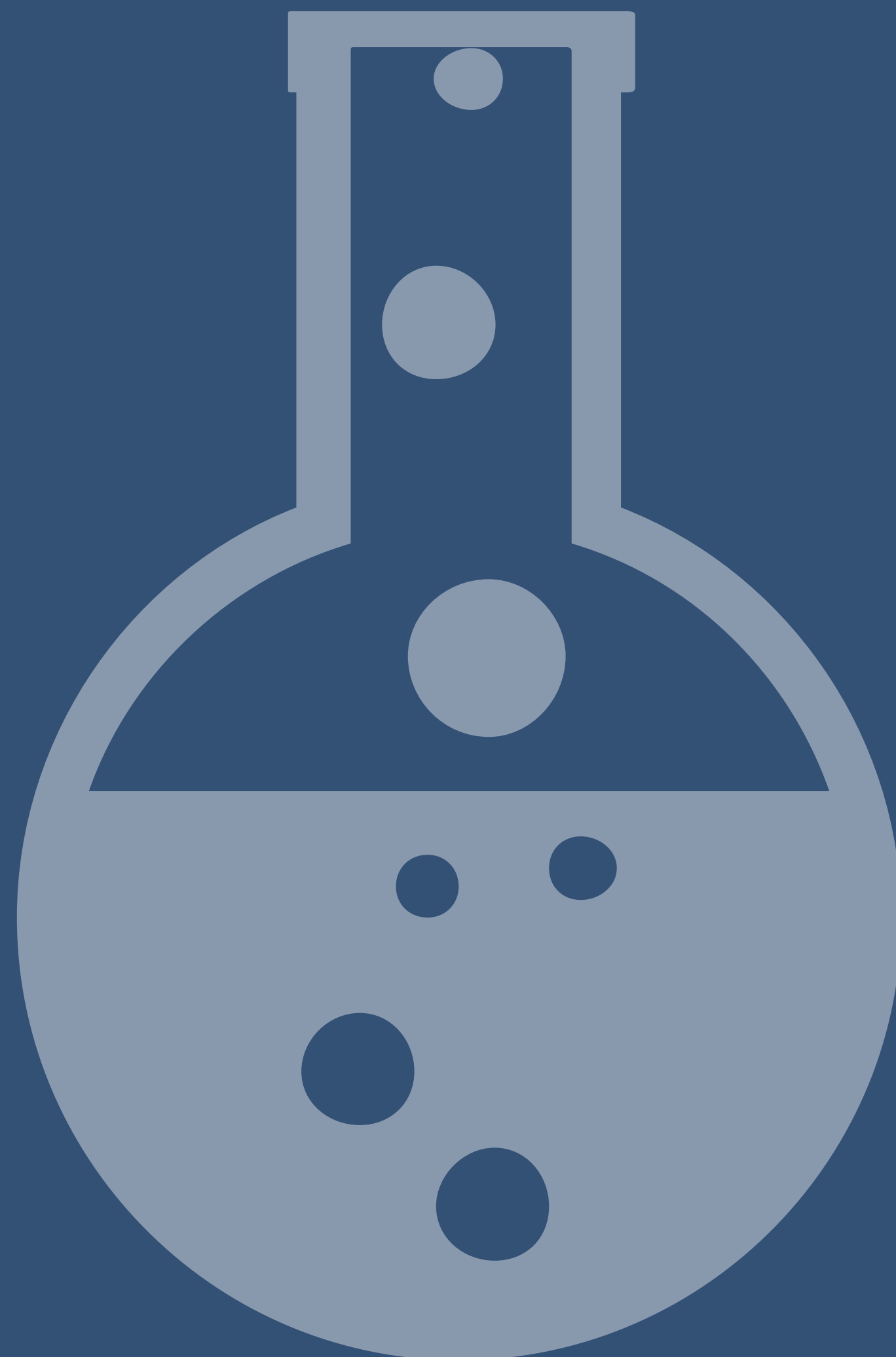
03

IT IS NOT ALWAYS AS IT SEEMS

Some questions may touch on topics that hint at underlying issues not explicitly stated, highlighting the need for an interdisciplinary approach to fully address them.

Future work

Next steps in understanding hidden dynamics.



More about the unspoken

In approx. 12% of cases, users submit a follow-up question. Starting in 2022, these subquestions are coded separately. We are particularly interested in identifying the topics that emerge in these follow-ups.

Gender, age, and topics

We aim to explore which topic co-occurrences are strongly influenced by gender and age. Can gender and age help predict the unspoken questions?