Topological methods in computational ecology

Session 8: Mathematical and statistical models to predict and protect tropical species and ecosystems
8th European Conference of Tropical Ecology
February 27, 2025, Amsterdam

Jānis Lazovskis

UNIVERSITY OF LATVIA

Juliano Morimoto

1495 UNIVERSITY OF ABERDEEN

Ran Levi

1495 UNIVERSITY OF ABERDEEN

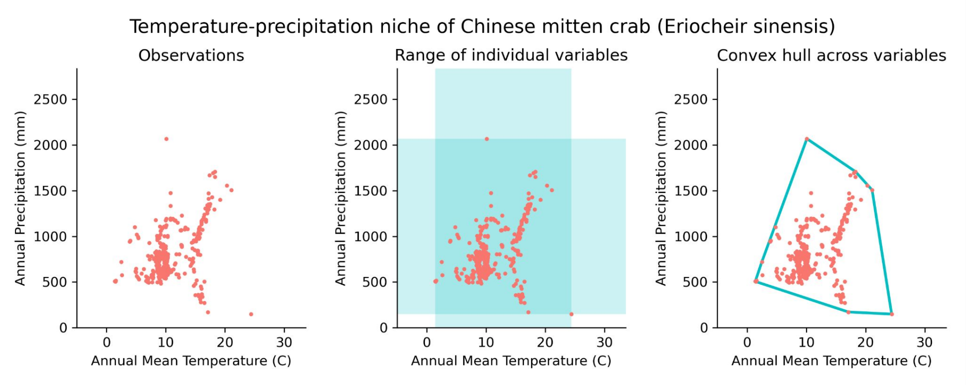
The hypervolume as an **ecological concept**

Given a species, its **niche** (Hutchinson, 1957) is the set of environmental parameters in which it can exist.

- The size of a climatic niche hypervolume is hypothesized to drive species *diversification rates*
- The similarity of species' environmental or functional trait hypervolumes measures *niche divergence* or packing, which may influence species *coexistence and richness patterns*
- Niche similarity also helps compare individuals within a species, assessing *climate change impacts* and niche shifts during *invasions*

The **realized niche** is the subset of the fundamental niche that is observed.

Hypervolumes were expected to be **convex**, but this assumption has been contested. Hard to unequivocally answer, as all animals cannot be observed, so **inferences** have to be made.



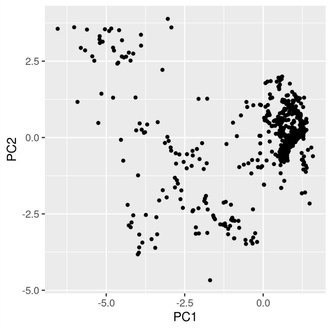
What is the smallest, most reasonable space, in which observations are made?

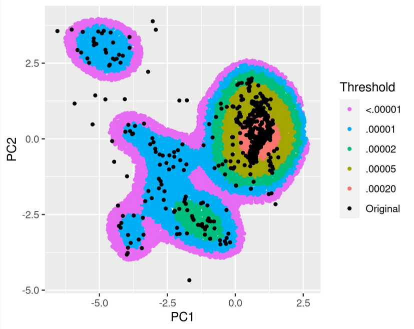
The hypervolume as a **computational object**

GBIF → species observations → (normalized) climate data at observations → project to lower-dimensional space (e.g. PCA) → compute hypervolume

GBIF
Global Biodiversity Information Facility

```
> data_raw %>% head
  ID alt bio1 bio10 bio11 bio12 bio13 bio14 bio15 bio16 bio17 bio18 bio19 bio2
1  1  30  144   185    99   932   206     1    93   521     9    13   521  101
2  2  48  140   187    91   753   168     1    92   423     9    13   423  138
3  3   4  155   219    84   361    80     1    85   197     6     9   197  137
4  4  10  144   191    92   740   174     1    95   430     7    10   430  128
5  5   0  150   194   102   356    73     1    86   186     5     8   185  108
6  6   2  152   211    87   518   119     1    90   287     6    10   287  134
  bio3 bio4 bio5 bio6 bio7 bio8 bio9         x        y dcoast
1   52 3393  244   50  194   99  182 -122.4583 37.95833     -3
2   57 3812  273   33  240   91  183 -122.6250 38.20833     20
3   47 5333  315   27  288   84  218 -121.7917 38.04167     42
4   54 3936  270   37  233   92  188 -122.4583 38.12500      5
5   52 3613  253   49  204  111  191 -122.0417 37.45833      3
6   49 4891  299   30  269   87  209 -122.0417 38.20833     28
>
```





Holes in a hypervolume suggest the opportunity for another species to take over, or an impending extinction.

↓

We want to find holes!

Problem: Gaussian KDE blurs each observation and softens jagged features = potential holes

↓

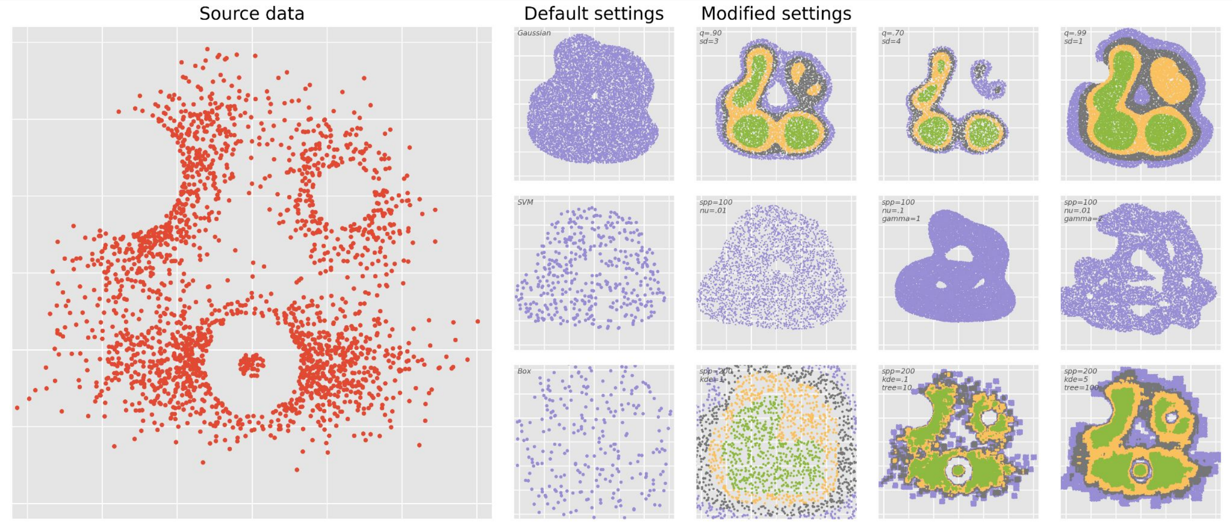
Construction parameters can vary

Solution: Infer unobserved data by averaging tight collections of observed data

↓

Blur at short distances, respect all data

Example: holes and non-convex features

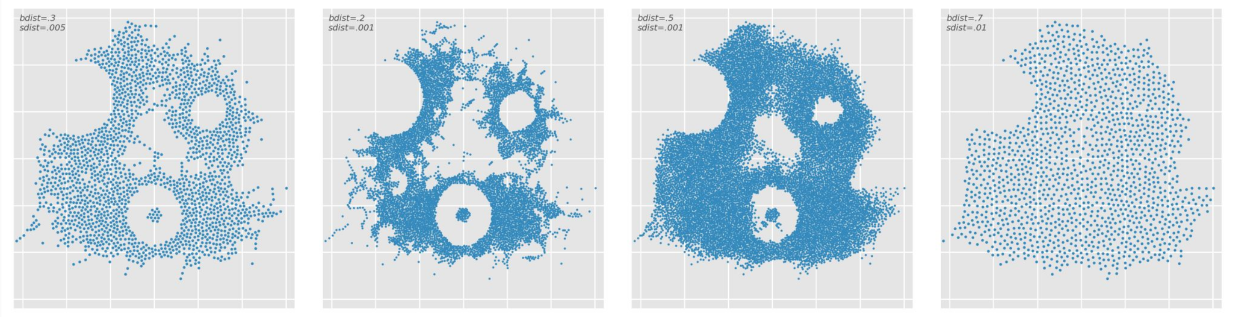


hypervolume: High Dimensional Geometry, Set Operations, Projection, and Inference Using Kernel Density Estimation, Support Vector Machines, and Convex Hulls

Estimates the shape and volume of high-dimensional datasets and performs set operations: intersection / overlap, union, unique components, inclusion test, and hole detection. Uses stochastic geometry approach to high-dimensional kernel density estimation, support vector machine delineation, and convex hull generation. Applications include modeling trait and niche hypervolumes and species distribution modeling.

Version:	3.1.5
Depends:	Rcpp, methods, R (≥ 3.5.0)
Imports:	raster, maps, MASS, geometry, ks, hitandrun, pdist, fastcluster, compiler, e1071, progress, mvtnorm, data.table, terra, sp, foreach, doParallel, parallel, ggplot2, pbapply, palmerpenguins, purrr, dplyr, caret
LinkingTo:	Rcpp, RcppArmadillo, progress
Suggests:	rgl, magick, alphahull, knitr, rmarkdown, gridExtra
Published:	2025-01-17
DOI:	10.32614/CRAN.package.hypervolume
Author:	Benjamin Blonder [aut, cre], Cecina Babich Morrow [aut], Stuart Brown [aut], Gregoire Butruille [aut], Daniel Chen [aut], Alex Laini [aut], David J. Harris [aut], Clement Violet [aut]
Maintainer:	Benjamin Blonder <benjamin.blonder at berkeley.edu>
BugReports:	https://github.com/bblonder/hypervolume/issues
License:	GPL-3
URL:	https://github.com/bblonder/hypervolume
NeedsCompilation:	yes
CRAN checks:	hypervolume results

R package “hypervolume”



C++ headers “TopoAware”

https://github.com/jlazovskis/TopoAware

README GPL-3.0 license

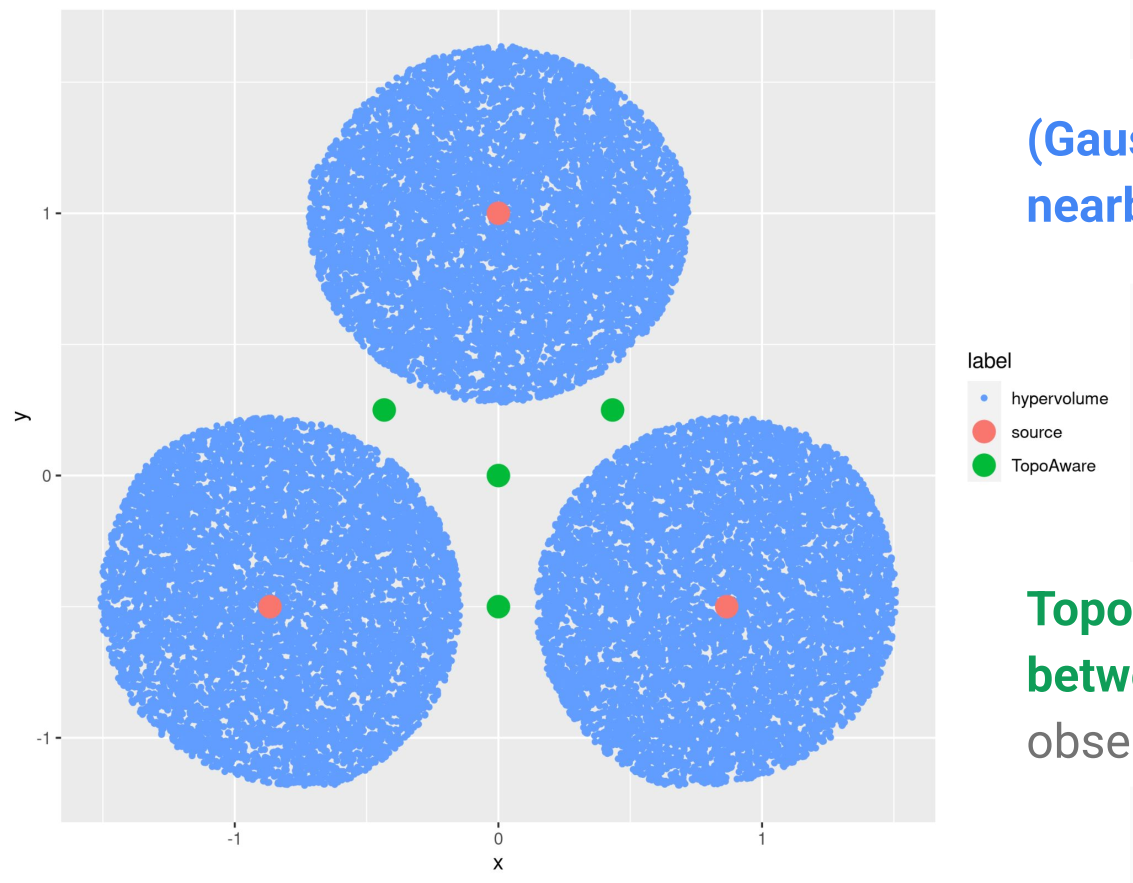
TopoAware

Topologically aware constructions for ecological hypervolumes

About

The purpose of this software is to use topological tools, in terms of computational efficiency and theoretical guarantees, to construct hypervolumes for use in ecology. Hypervolumes are usually constructed as kernel density estimators, but those are often less interesting topologically at the expense of knowing more information (precisley "filling in" the holes of missing data). This software aims to retain toological information of the input sample while still providing the user with more information about the space in which the sample lies.

Extreme example: complementary methods



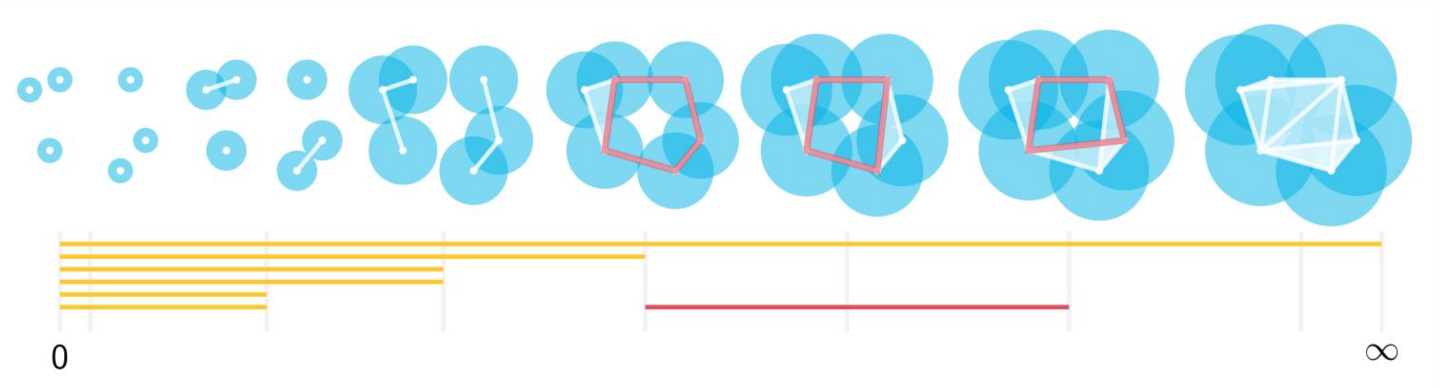
(Gaussian) KDE infers new data **nearby** each true observation

Topology infers new data **between** collections of true observations

Topology and persistent homology

Topology is the mathematics of surfaces and shapes, focusing on classifying spaces by their fundamental, immutable features. Features are components, holes, voids, etc.

Topological data analysis is the application of topology to datasets, usually by recognizing the features of the topological spaces associated to datasets.



connected components are 0-dimensional features

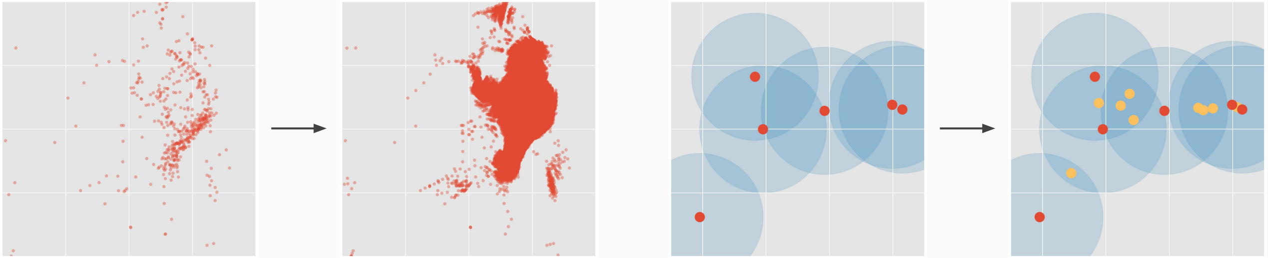
holes are 1-dimensional features

Persistent homology (PH) is a tool in topological data analysis that observes significant features in a dataset, from small to large measures of "nearness". A stable method, robust under noise and distortion.

- The output of PH is a collection of *intervals* indicating the *birth* and *death* of features
- PH has been applied to a wide range of scientific work since its inception in the early 2000's

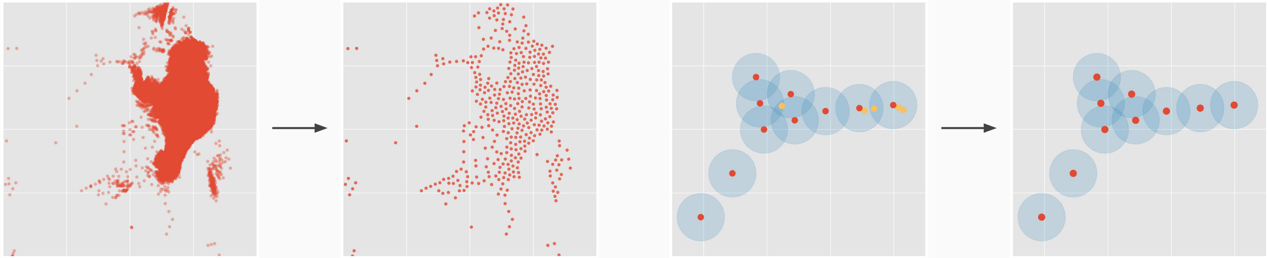
TopoAware: Topologically aware constructions for ecological hypervolumes

Barycentric subdivision: Add average of pairs and triples, whenever less than distance d_b



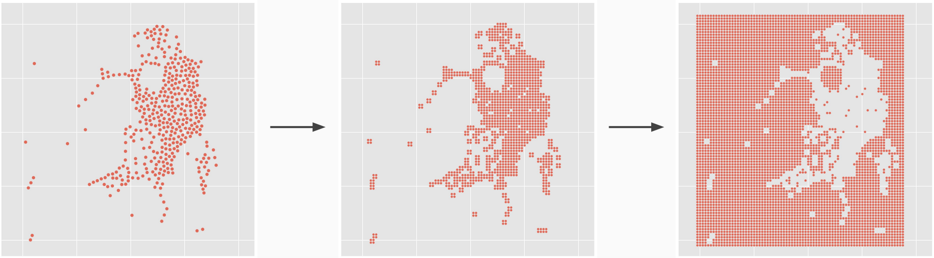
Infer new data

Sparsification: Remove all points ordered after and within chosen distance $d_s < d_b$ of each point



Distribute evenly

Alignment to grid / complement: Use known grid size to simplify computations



Components of complement are features in the highest dimension

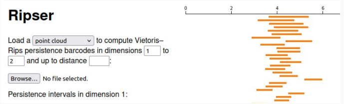
Arrange for computation

Use cases

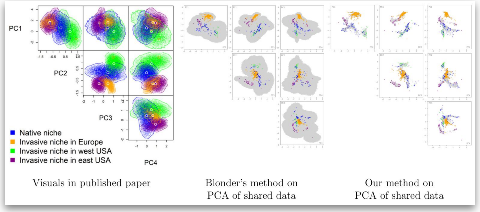
TopoAware is a tool for supporting *analysis of large data sets in ecology*.

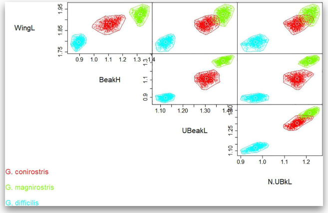
- **Holes:** Count how many holes are in a hypervolume, how big each hole is
    - Constructs inputs for persistent homology software, for computing holes and their signfcance

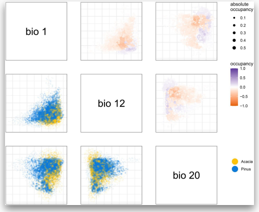
गुढी GUDHI Geometry Understanding in Higher Dimensions

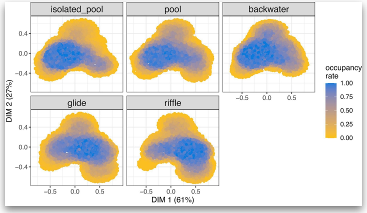


- **Validation:** Use TopoAware in place of similar methods to generate and analyze hypervolumes
    - Additional support or contesting evidence; replicability of findings



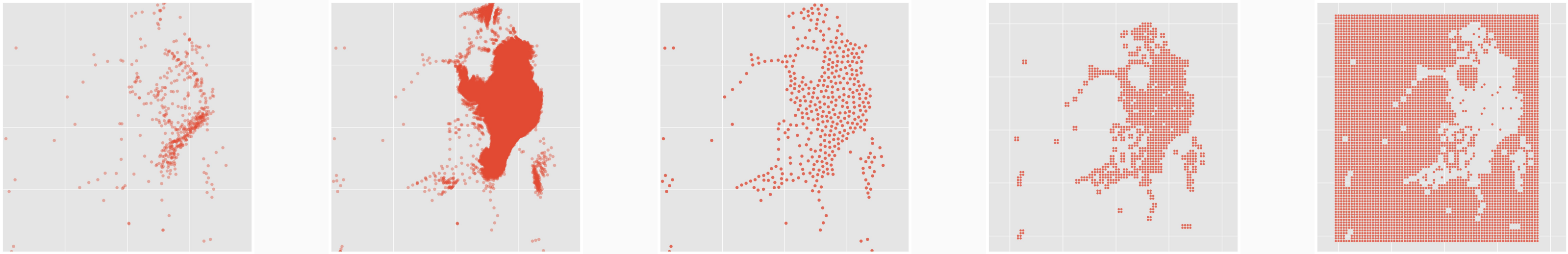






- **Data wrangling:** Uniformize data sets, decrease the size, while keeping fundamental features
    - Program is modular, with dataframes as input / output for each operation

Still in the development stage - suggestions and comments welcome!



- Hutchinson, G Evelyn. **Population studies-animal ecology and demography - concluding remarks**. Cold Spring Harbor symposia on quantitative biology, 1957.
- Carlsson, Gunnar. **Topology and Data**. Bulletin of the American Mathematical Society, 2009.
- Blonder, Benjamin, et al. **The n-dimensional hypervolume**. Global Ecology and Biogeography, 2014.
- Blonder, Benjamin. **Do Hypervolumes Have Holes?** The American Naturalist, 2016.
- Blonder, Benjamin, et al. **New approaches for delineating n-dimensional hypervolumes**. Methods in Ecology and Evolution, 2018.
- Eddelbuettel, Dirk and James Joseph Balamuta. **Extending R with C++: A Brief Introduction to Rcpp**. The American Statistician, 2018.
- Zhang, Zhixin, et al. **To invade or not to invade? Exploring the niche-based processes underlying the failure of a biological invasion using the invasive Chinese mitten crab**. Science of The Total Environment, 2020.
- Carvalho, José Carlos and Pedro Cardoso. **Decomposing the Causes for Niche Differentiation Between Species Using Hypervolumes**. Frontiers in Ecology and Evolution, 2020
- Bauer, Ulrich. **Ripser: efficient computation of Vietoris-Rips persistence barcodes**. Journal of Applied and Computational Topology, 2021.
- Giunti, Barbara, Jānis Lazovskis, and Bastian Rieck. **DONUT: Database of Original & Non-Theoretical Uses of Topology**. Online database, 2022.
- Laini, Alex, Thibault Datry, and Benjamin Wong Blonder. **N-dimensional hypervolumes in trait-based ecology: Does occupancy rate matter?** Functional Ecology, 2023.
- Lazovskis, Jānis. **TopoAware: Topologically aware constructions for ecological hypervolumes**. GitHub, 2024.
- Chen, Daniel, Alex Laini, and Benjamin Wong Blonder. **Statistical inference methods for n-dimensional hypervolumes: Applications to niches and functional diversity**. Methods in Ecology and Evolution, 2024.
- Jamin, Clement, Siargey Kachanovich, and Marc Glisse. **GUDHI subsampling**. GUDHI Editorial Board, 2025.

Thank you for your attention

slides available at jlazovskis.com/talks