

Data. Analytics. Dissemination.

Wolfgang Karl Härdle
Xiaorui ZUO

Ladislaus von Bortkiewicz Professor of Statistics
Humboldt-Universität zu Berlin
Fudan University, Shanghai
theIDA.net

ASE, ACI NUS, NYCU 玉山學者

Platograph - D.A.D. - Q2 Ecosystem

AIoT/ESG – Sensor Network
圖靈尤拉計畫系列研討會 II.3

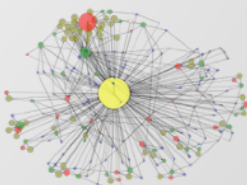
- 感測器網絡AI建模關鍵架構技術研討
- 企業AI升級轉型家族傳承實驗班
- TUMC AoI企業數位大腦展演
- TUMC企業畢業季聯合徵才
- 工業4.0三部曲 – IA-II-AI

TuringEuler Project

北科大先鋒國際研發大樓
 06/29/2024 2-5pm

主辦單位：TUMC 協辦單位：台灣工商總會
 國立台北科技大學 柏林洪堡大學 Q2

杜風工程服務股份有限公司
 磊鑫國際企業有限公司
 協得科技股份有限公司
 恆析家族辦公室（香港）有限公司



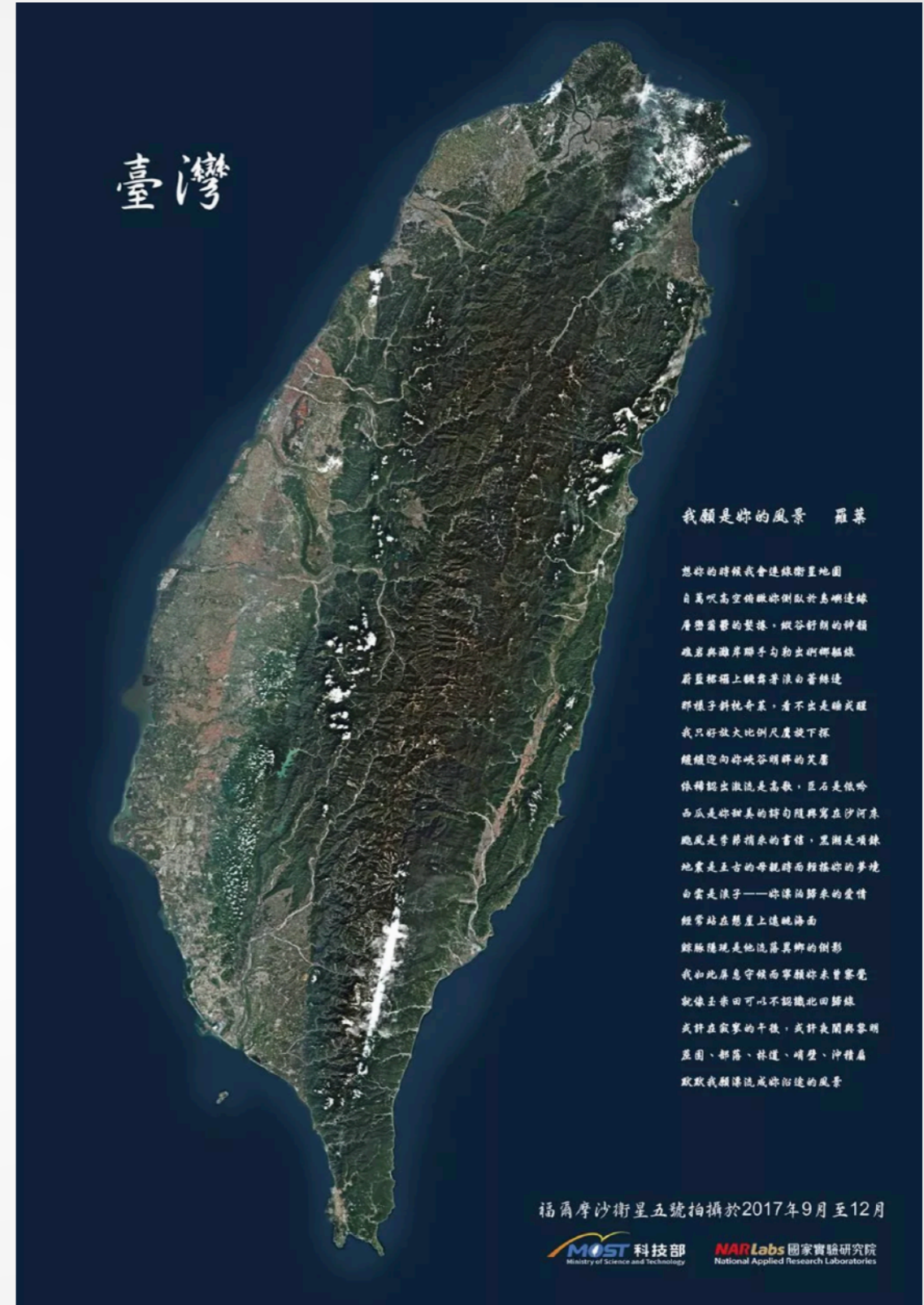
Formosat-5

- ▣ Remote Sensing
- ▣ Vegetation
- ▣ Sea Level
- ▣ Climate Control
- ▣ Sea Surface Temperature

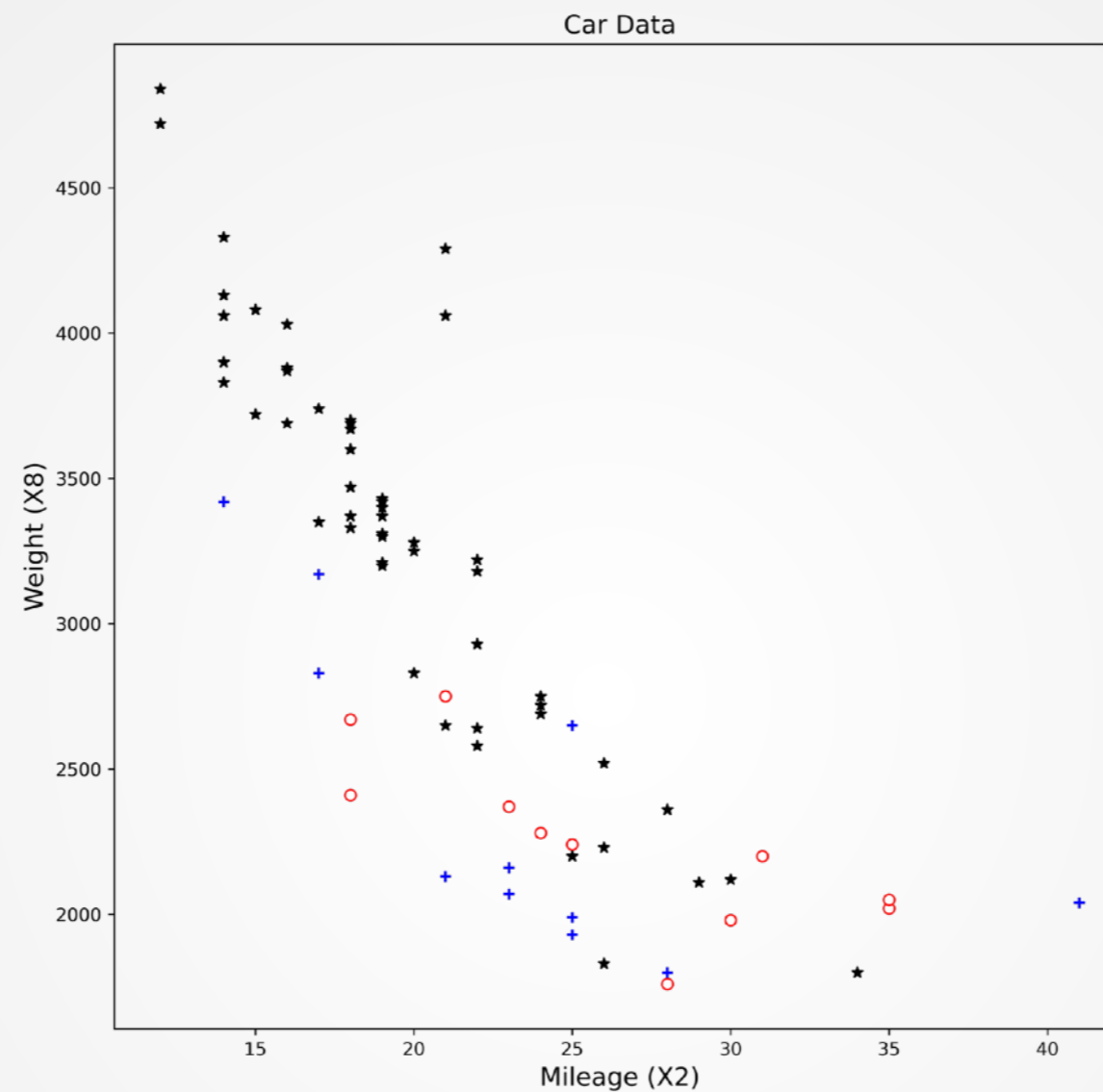
A Taiwanese Satellite

Formosat-5 Highlights Space R&D Capabilities

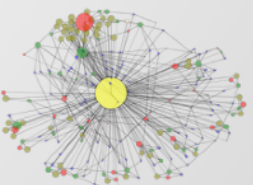
Esther Tseng / photos NSOP / tr. by Scott Williams
April 2018



Carbon Emissions

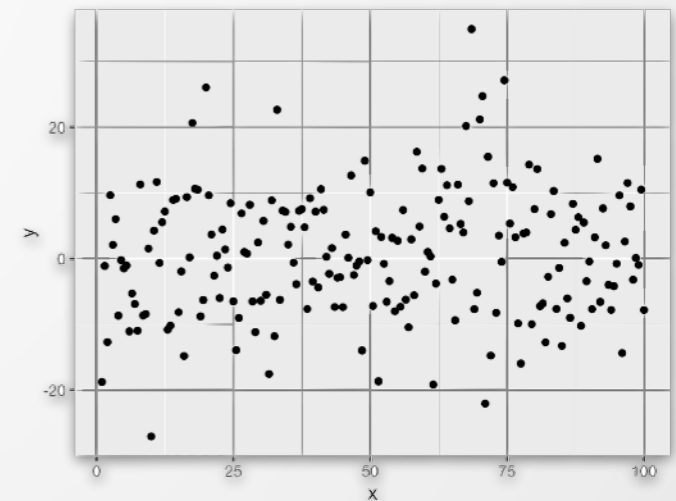
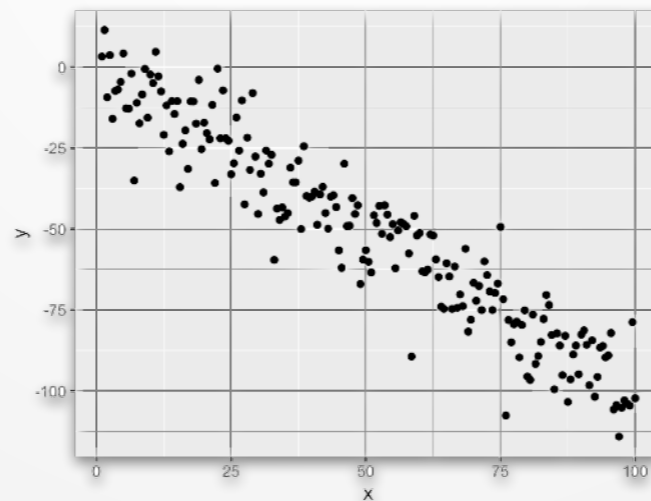
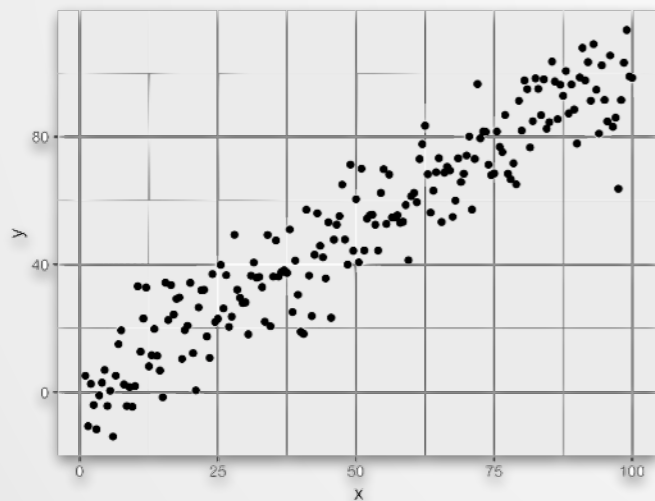


Mileage (X_2) vs. weight (X_8) of U.S. (\star), European ($+$) and Japanese (\circ) cars.

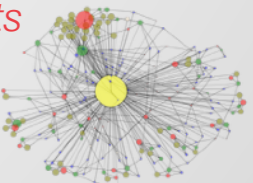


GPT based ANOVA

- Scatterplots with point clouds that are "upward-sloping" are showing variables with positive covariance.
- Scatterplots with "downward-sloping" structure have negative covariance.
- Scatterplots with „no sloping“ shape indicate zero covariance.



Created with Chat GPT on 20230119:
an upwards sloping scatterplot with 100 data points



Example: car data set

Correlation between mileage (X_2) and weight (X_8)

$$n = 74, r_{X_2X_8} = -0.823$$

$$H_0 : \rho = 0 \quad H_1 : \rho \neq 0$$

$$w = \frac{1}{2} \log \left(\frac{1 + r_{X_2X_8}}{1 - r_{X_2X_8}} \right) = -1.166 \quad , \quad z = \frac{-1.166 - 0}{\sqrt{\frac{1}{71}}} = -9.825$$

$$H_0 : \rho = -0.75$$


$$z = \frac{-1.166 - (-0.973)}{\sqrt{\frac{1}{71}}} = -1.627.$$



ANOVA

- Is the E - Porsche a good deal?

Moving to Higher Dimensions
3-55




ANOVA

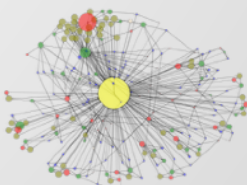
 Table

SS	df	MS	F-stat	p-value
SS(explained)	$p - 1$	$\frac{ss(\text{explained})}{p-1}$	$\frac{ss(\text{explained})/(p-1)}{MSE}$	p-value
SS(full)	$n - p$	$\frac{ss(\text{full})}{n-p} = MSE$		
SS(reduced)	$n - 1$			

$F \sim F_{p-1, n-p}$

Test: reject H_0 if $F > F_{1-\alpha; p-1, n-p}$, or if p-value $< \alpha$

Applied Multivariate Statistical Analysis




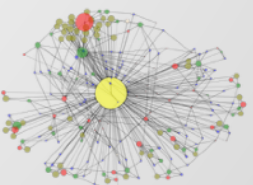
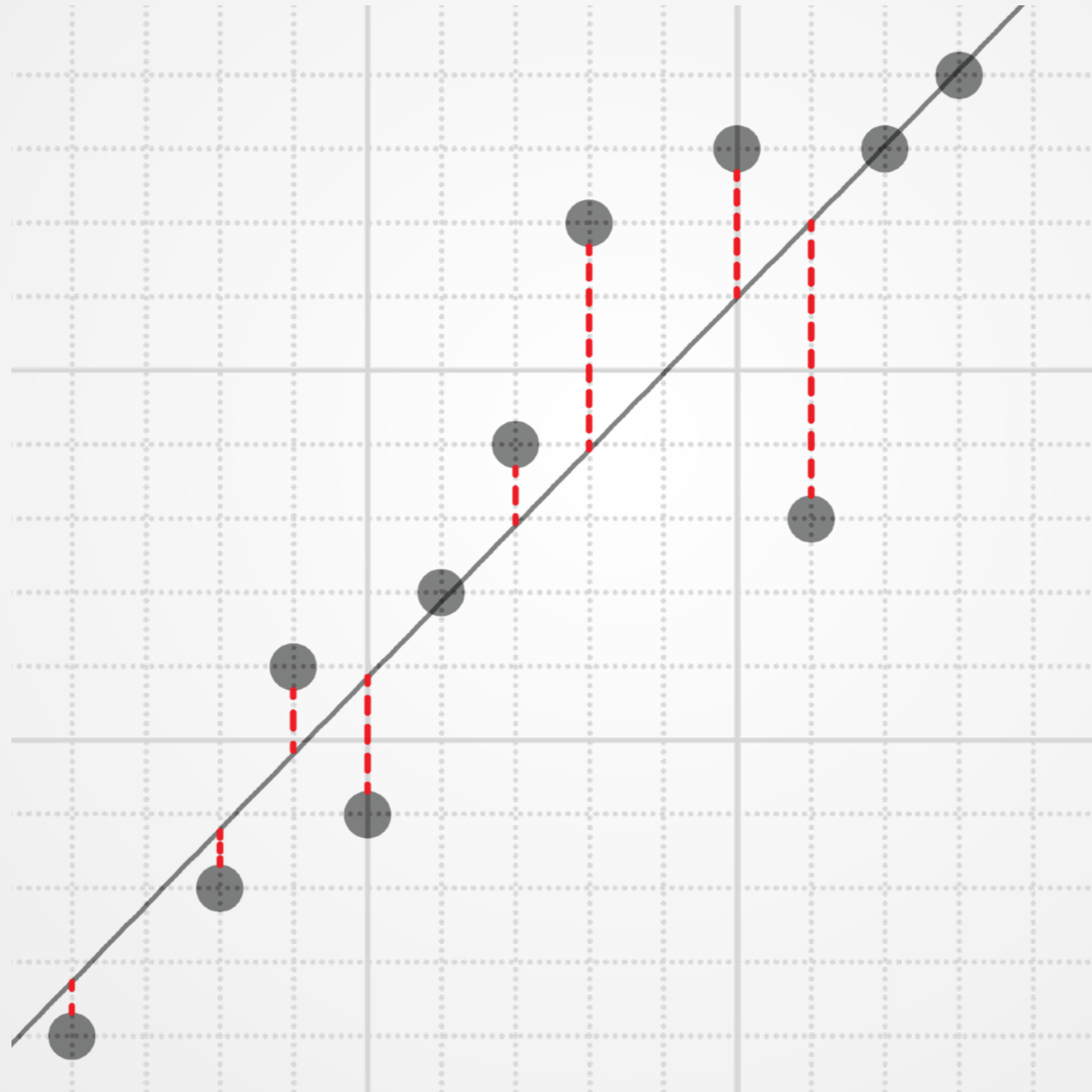
Outline

1. Motivation ✓
2. ANOVA & Remote Sensing
3. D. A. D.



One ANOVA slide only !

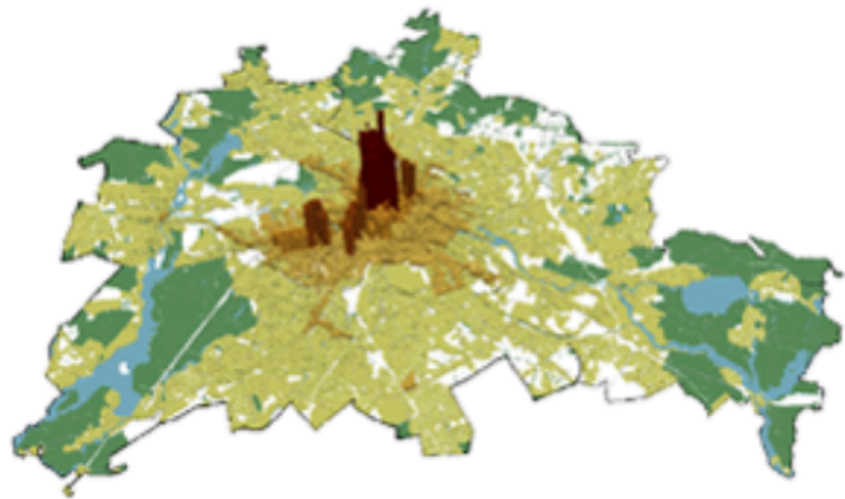
- Sum the (red dashed lines)², compare with a constant



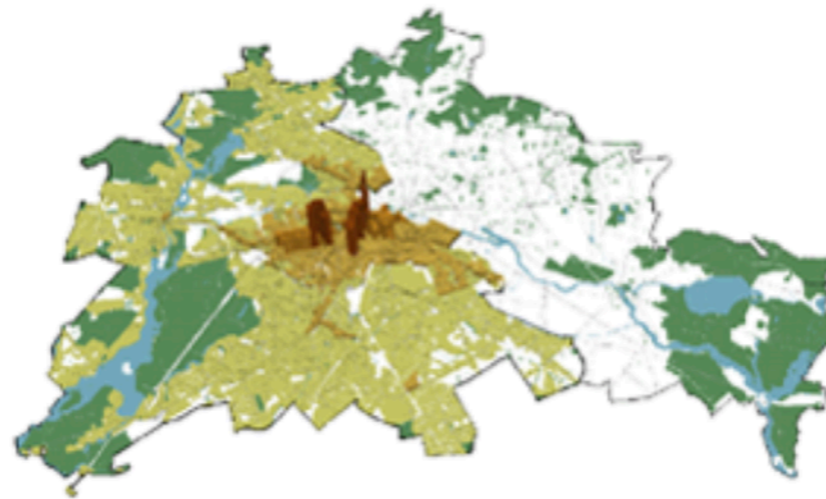
Spatial Monitoring

- Evolution of land prices in Berlin

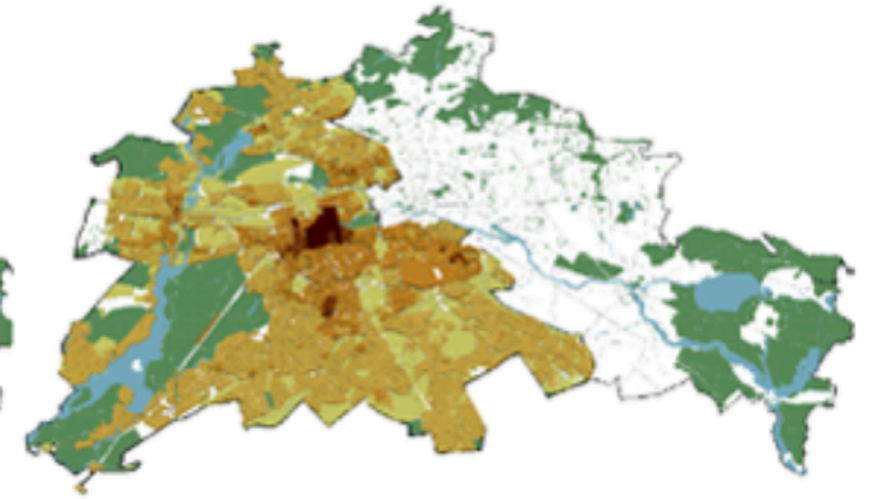
Panel A: Berlin 1936



Panel B: West Berlin 1936



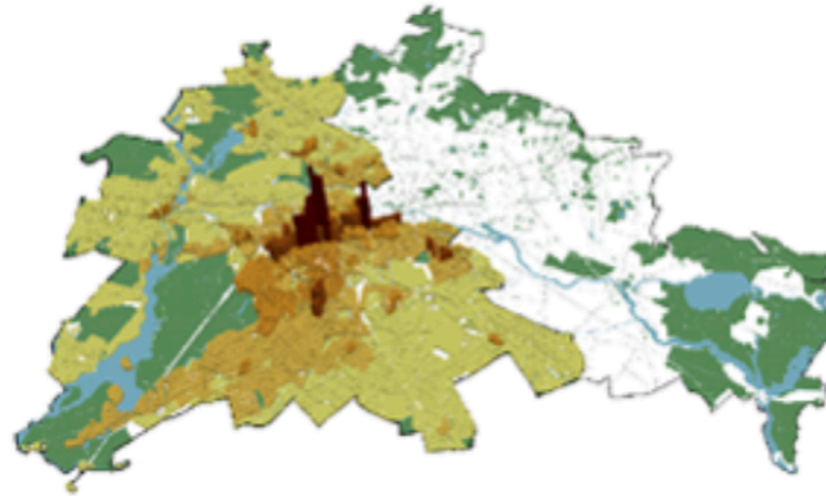
Panel C: West Berlin 1986



Panel D: Berlin 2006



Panel E: West Berlin 2006

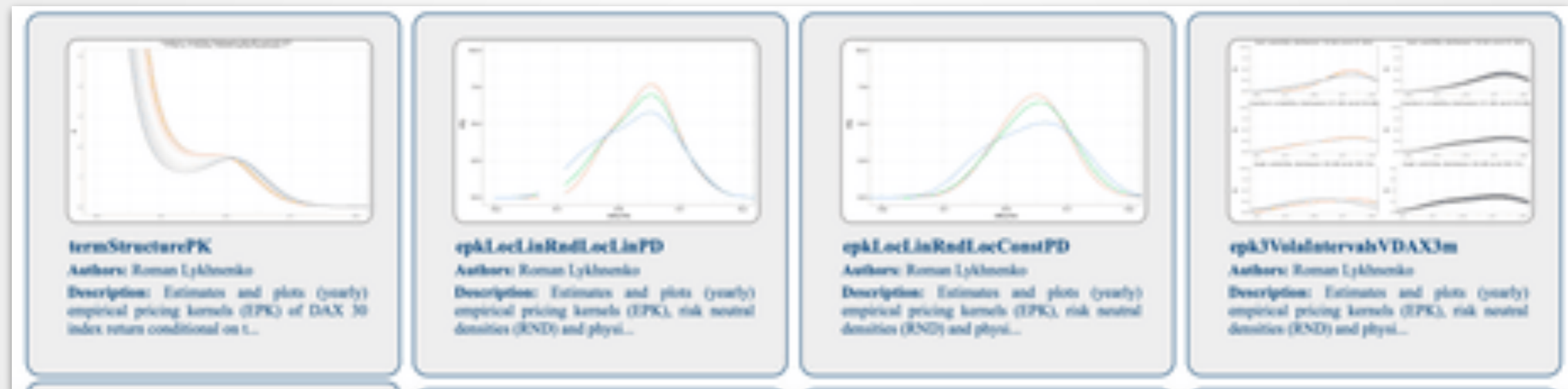


The Economics of Density: Evidence From the Berlin Wall



Functional Data Analytics

- Remote Sensing Data
- Spatial dependencies
- Predict future curves



Data and information are everywhere ...

INTRO
 PROF ERNEST BREHMER, SMU

WELCOME
 PROF GERARD GEORGE, SMU

KEYNOTE
 DR DAVID KUI HARBOON, MAS

CONTEXTUALISE DATA SCIENCE
 MACHINE-LEARNING
 SHARE INSIGHTS
 COLLECTIVELY
 WHEN IN DOUBT ASK
 KNOWLEDGE & EMBRACE CHANGE
 DATA SHOULD BE LINKED TO IDENTITY

BIG DATA APPLICATIONS
 FINANCIAL INCLUSION
 THE WORLD DATA IS everchanging
 BIG DATA ↑ SCOPE GRANULARITY
 RECOGNIZE CHALLENGES & OPPORTUNITIES
 DIGITAL TRANSFORMATION
 TRACK DIGITAL & PHYSICAL FOOTPRINT
 BEHAVIOUR PREDICTION
 MERCHANT AFFINITY
 STRONG & RESEARCH
 ENHANCE SERVICE
 SMART HIRING & TRAINING

SMART DATA ANALYTICS
 CHALLENGES INTO PROMISES
 IDENTITIES → ABILITY TO BUILD CUSTOMER LOYALTY
 INSIGHTS → USE THE DATA (EG. CREDIT SCORES) TO EXPAND BANKING SERVICES
 PITFALLS → WHAT EVERYTHING IS MEASURABLE, REASONABLE OR DESIRABLE!
 SOCIAL INFRASTRUCTURE → EXCLUSIONS & ABUSE OF POWER

ADVANCES IN DATA SCIENCE & IMPLICATIONS FOR BUSINESS
 3RD ANNUAL SKBI CONFERENCE 2017

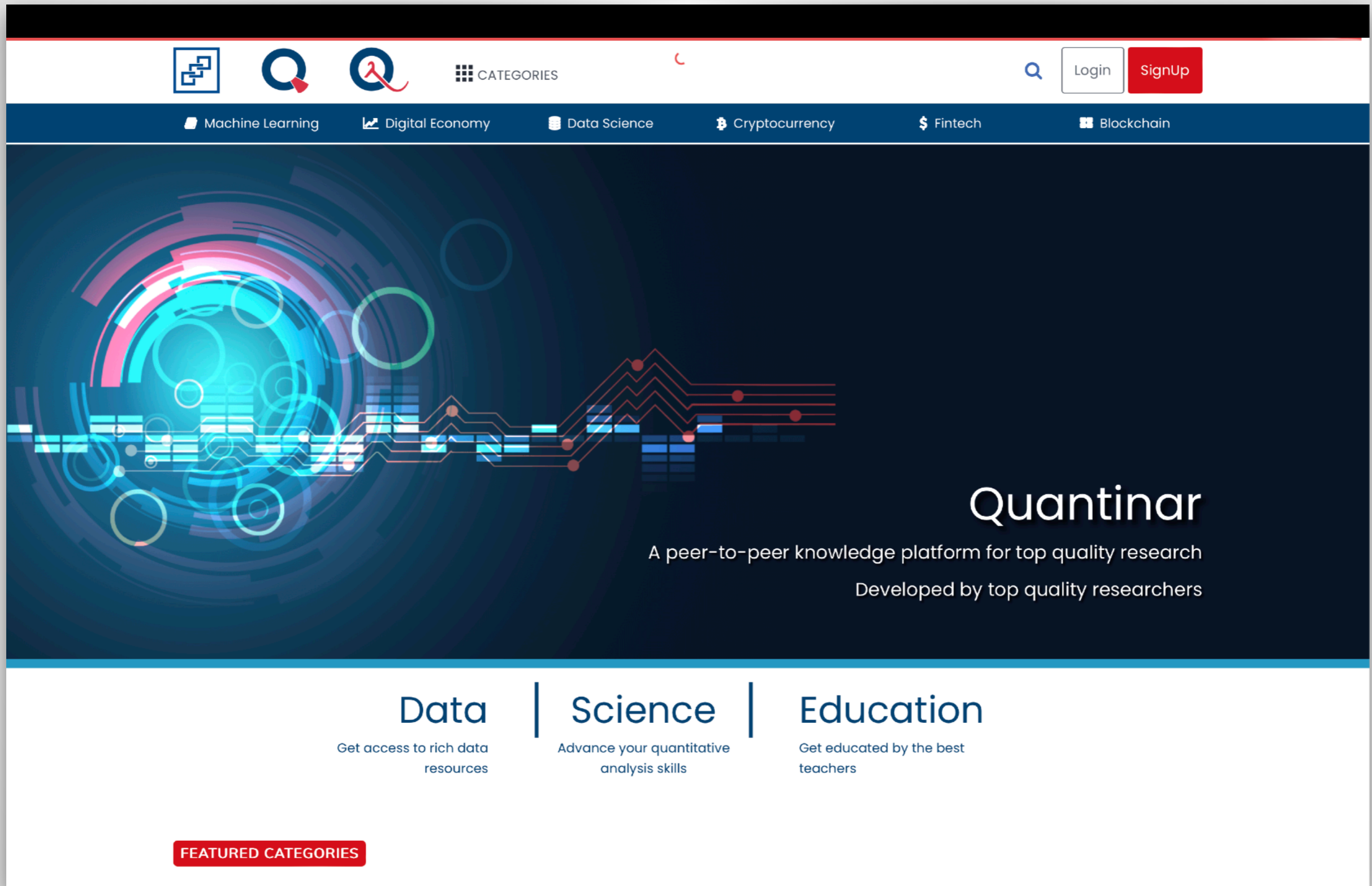
SMART DATA ANALYTICS
 UNLOCK DARK DATA INTO SMART DATA
 CRYPTOCURRENCY BECOMES MAINSTREAM
 PREDICT INDUSTRY TRENDS, ASSET PRICES ETC

EXPERIENCES WITH DATA @ START-UPS
 VALUE PROPOSITION
 HOW WE USE DATA
 HOW WE DIFFERENTIATE FROM COMPETITORS

CLOSING
 CLASS VARIABLE TRANSFORMATION

idea ink WWW.IDEAINK.CO
 IDEA INK ILLUSTRATIONS @IDEAINK







The image shows the top portion of the Quantinar website. At the top is a white navigation bar with several icons: a blue square with a white staircase, a blue circle with a white magnifying glass, a blue circle with a white line graph, and a red circle with a white line graph. To the right of these icons is the word "CATEGORIES" in blue, followed by a red search icon, a white "Login" button, and a red "SignUp" button. Below the navigation bar is a dark blue horizontal bar with white text and icons for "Machine Learning", "Digital Economy", "Data Science", "Cryptocurrency", "Fintech", and "Blockchain". The main hero section has a dark blue background with a complex, colorful graphic of overlapping circles and lines in shades of blue, pink, and green. On the right side of this section, the word "Quantinar" is written in a large, white, sans-serif font. Below it, in a smaller white font, is the text "A peer-to-peer knowledge platform for top quality research" and "Developed by top quality researchers". At the bottom of the hero section is a white horizontal bar with three columns. Each column has a large blue heading: "Data", "Science", and "Education". Below each heading is a smaller blue sub-heading: "Get access to rich data resources", "Advance your quantitative analysis skills", and "Get educated by the best teachers". At the bottom left of this white bar is a red button with the white text "FEATURED CATEGORIES".



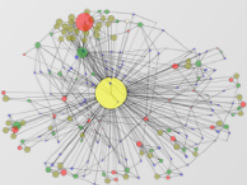
Q2 Quantinar & Quantlet Ecosystem

- ▣ Transparency and reproducibility in research and education
- ▣ Interaction with data, algorithms and community members
- ▣ ode and data with uantlets
- ▣ Open peer review; rewards, certificates
- ▣ P2P/TikTok Style

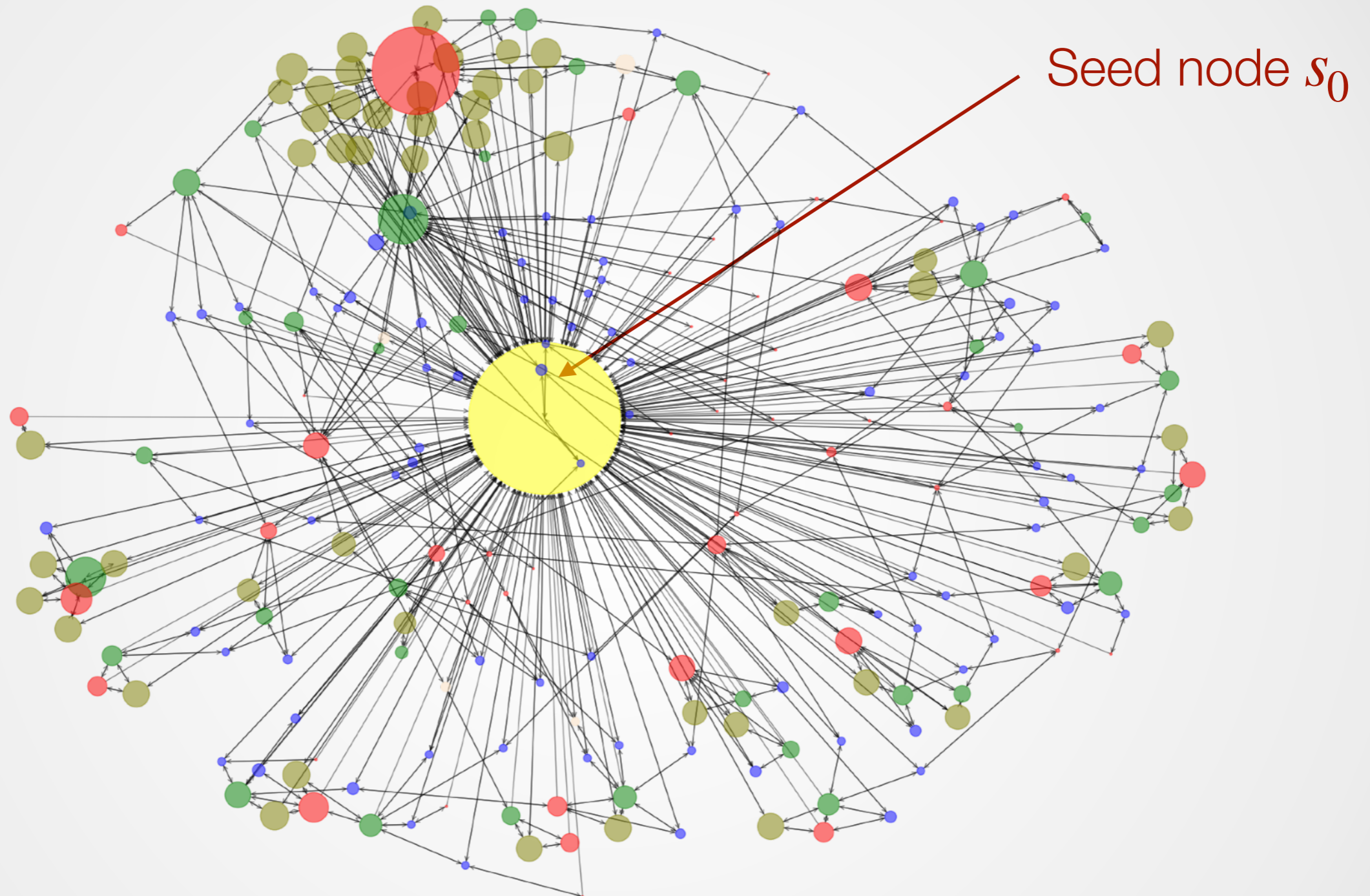


In my humble opinion, OCW is a good place to find class notes for some specific topics and/ or difficult problem sets to work on once you have a rudimentary understanding of the subject whereas Coursera or Edx should be used if you want to cover **the course in its entirety.**

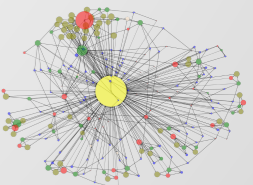
<https://www.quora.com/What-is-the-difference-between-MIT-OCW-and-Coursera-or-edx>



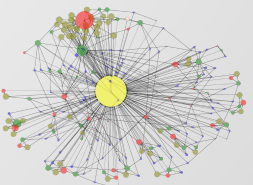
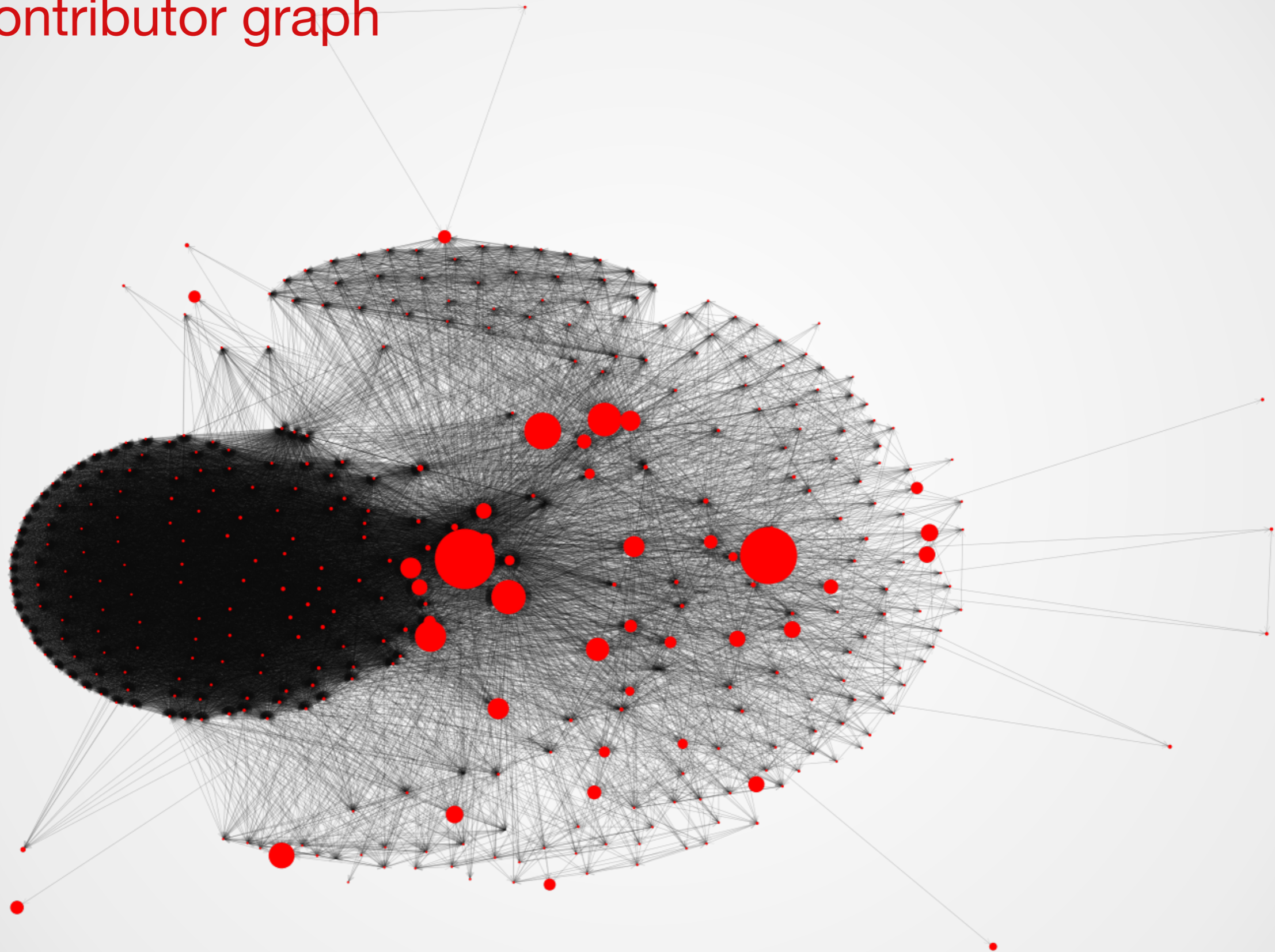
Contribution graph



PageRank Contribution graph at 2022-10-19: **seed**,
contributor, **course**, **courselet**, **order**, **review**

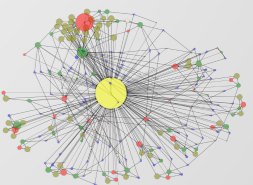


Contributor graph

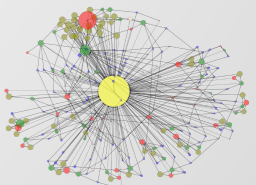
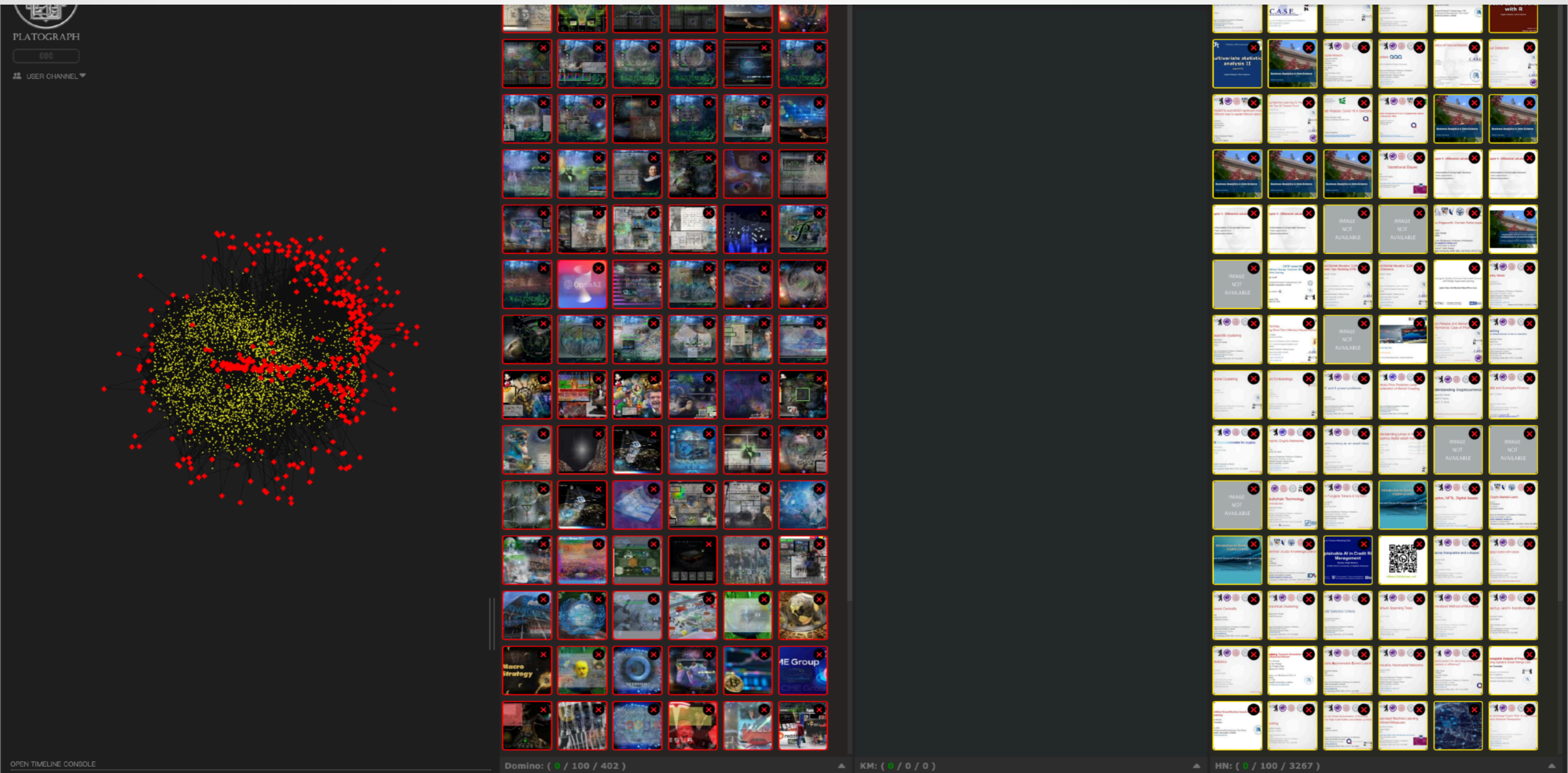




PLATOGRAPH



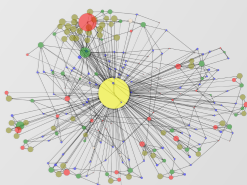
Platograph



ANOVA

The screenshot displays the Platograph data analysis interface. At the top, there is a search bar containing the text "blockchair" and a filter set to "WIDE/ACCURATE". Below the search bar are navigation buttons: "BLOOM", "SHOW ALL", "CLEAR", and "CLEAR ALL". A "sort" dropdown menu is set to "Collect Date". The main workspace is divided into three sections: a large network graph on the left, a central grid of image thumbnails, and a right-hand panel with a context menu. The network graph shows a complex web of nodes and edges, with a cluster of nodes highlighted in red. The image grid contains various thumbnails, some of which are labeled "IMAGE NOT AVAILABLE". The context menu on the right includes options like "buoy loaded", "buoy all", "buoy cancel all", "clear (keep buoyed)", "clear all", and "load all". At the bottom of the interface, there is a status bar with "Domino: (0 / 14 / 14)", "KM: (0 / 0 / 0)", and "HN: (0 / 46 / 46)". A bottom toolbar includes "INSERTION MODE: REPLACE" and a "WorkingSheet0" button.

Blockchain_mechanism_clustering – Kainat Khowaja , Min-Bin Lin, | Quantlet
– Advances in Quantitative Analysis of Finance & Accounting (AQFA) Use
distributional characteristics such as fourier power spectrum, moments,
quantiles, global we optimums, as well as the measures for long term
dependencies, risk and noise to summarise the information from crypto time
series and conduct clustering via spectral clustering Cryptocurrency,
Blockchain mechanism , Distributional characteristics,...

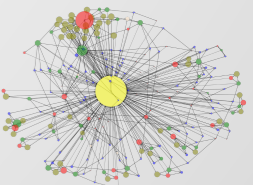


ANOVA

Source	SS	df	MS	F	Prob>F
Groups	102.6	2	51.3	8.78	0.0012
Error	157.7	27	5.8407		
Total	260.3	29			

SMSanovapull – Dana Chromikova, Awdesch Melzer | Quantlet – Multivariate Statistics – Exercises and Solutions performs a linear regression and an analysis of variance (ANOVA) for three marketing strategies. A company decides to compare the effect of three marketing strategies: 1. advertisement in local newspaper 2. presence of sales assistant 3. special presentation in shop windows, on the sales of their portfolio in 30 shops. The 30 shops were divided into 3 groups of 10 shops. The sales...

SMSanovapull – Dana Chromikova, Awdesch Melzer | Quantlet – Multivariate Statistics – Exercises and Solutions performs a linear regression and an analysis of variance (ANOVA) for three marketing strategies. A company decides to compare the effect of three marketing strategies: 1. advertisement in local newspaper 2. presence of sales assistant 3. special presentation in shop windows, on the sales of their portfolio in 30 shops. The 30 shops were divided into 3 groups of 10 shops. The sales...



quantlet.com

- Remote Sensing knowledge

Quantlet

kernel() Login Sign up

Portfolio HighDim Regression Time Series VotTech

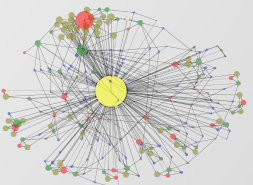
*Quantlet:
Where Knowledge Meets Code
fostering transparency, replicability, and
integrity in scientific exploration*

TermStructurePK
Authors: Roman Lykhenko
Description: Estimates and plots (yearly) empirical pricing kernels (EPK) of DAX 30 index returns conditional on t...

epkLocLinRadLocLinPD
Authors: Roman Lykhenko
Description: Estimates and plots (yearly) empirical pricing kernels (EPK), risk neutral densities (RND) and physi...

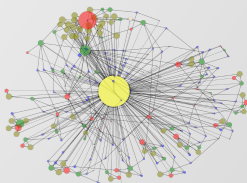
epkLocLinRadLocConstPD
Authors: Roman Lykhenko
Description: Estimates and plots (yearly) empirical pricing kernels (EPK), risk neutral densities (RND) and physi...

epk3VolIntervahVDAX3m
Authors: Roman Lykhenko
Description: Estimates and plots (yearly) empirical pricing kernels (EPK), risk neutral densities (RND) and physi...



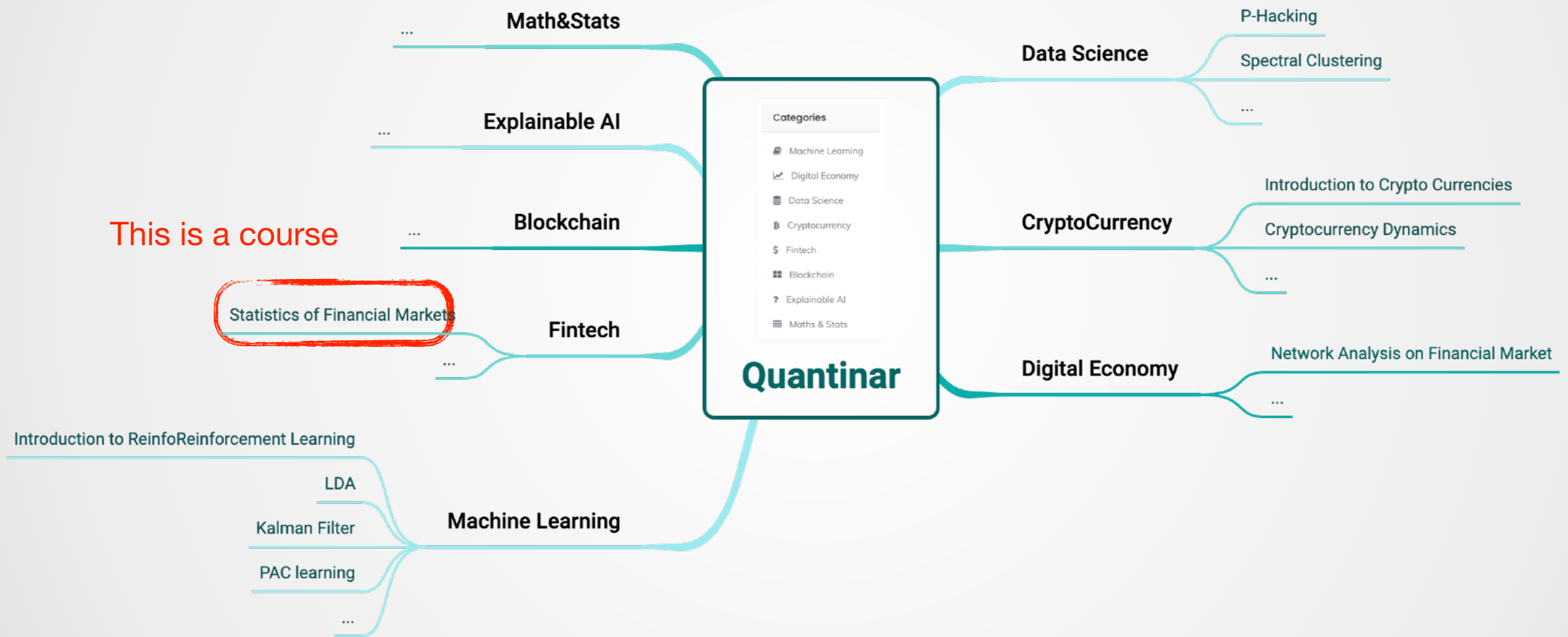
The screenshot shows the Quantlet website interface. At the top, there's a navigation bar with the Quantlet logo and search, login, and sign-up buttons. Below this is a dark blue header with category buttons: FinTech, HighDim, Regression, Time Series, and VizTech. The main content area features a large banner with the text: "Quantlet: Where Knowledge Meets Code fostering transparency, replicability, and integrity in scientific exploration". Below the banner, there are four project cards, each containing a plot and a description:

- SFM_Rough_Rice_Futures_Price**
Authors: Xingjia Wang
Description: (S)ARIMA Model for Rough Rice Futures Price from 2011 to 2021
- AdaBoost_P2PI**
Authors: Konstantin Häusler, Wolfgang Karl Härdle, Xiaorui ZUO
Description: AdaBoost classification model for defaults of P2P loans. Compares the out-of-sample predictive perfo...
- AdaptivePenalizedMacroFactor**
Authors: nan
Description: Builds a classification model to predict
- HermPolyPlot**
Authors: Sergey Nasekin, Katerina Papagiannouli
Description: Plots the first 4 Hermite polynomials on the given grid of

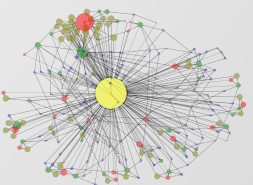


Recommender System

▣ Courses



Presented with xmind



Architecture of Quantinar

- Courselets

This is a courselet

Probability Theory

...

Statistics of Financial Markets

+ Chapter 1 - Financial Derivatives	2 Components 26 min
+ Chapter 2 - Basics of Option Management (1)	2 Components 24 min
+ Chapter 2 - Basics of Option Management (2)	2 Components 34 min
+ Chapter 2 - Basics of Option Management (3)	2 Components 26 min
+ Chapter 3 - Probability Theory	2 Components 27 min
+ Chapter 4 - Stochastic Processes in Discrete Time	2 Components 24 min
+ Chapter 5 - Stochastic Integrals and Differential Equations	2 Components 29 min
+ Courselet: Pricing Kernels	2 Components 47 min

- Chapter 1 - Financial Derivatives	2 Components 26 min
Chapter 1 - Financial Derivatives (video)	26 min
Chapter 1 - Financial Derivatives (pdf)	1.08 K

Financial Derivatives

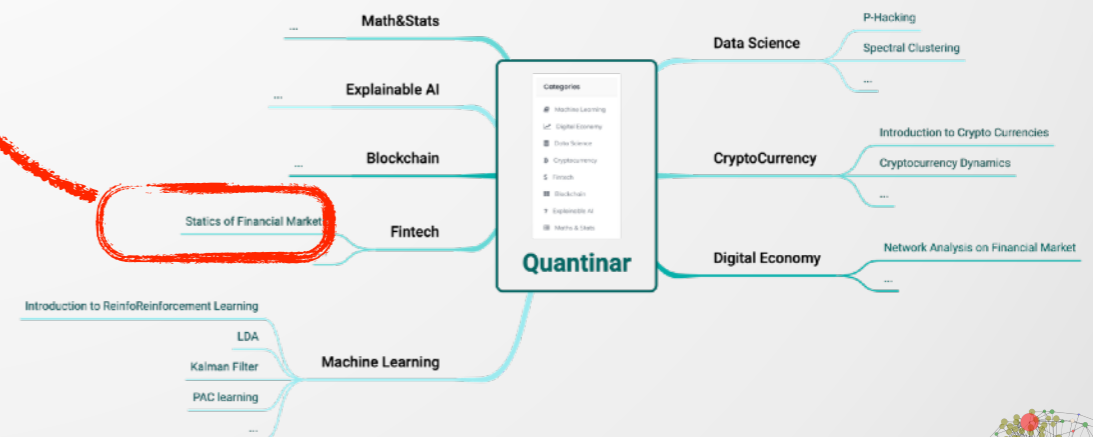
Video

pdf

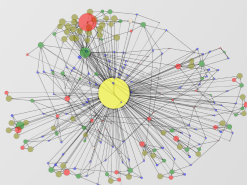
Those are resources

Basics of Option Management (1)

Presented with mind

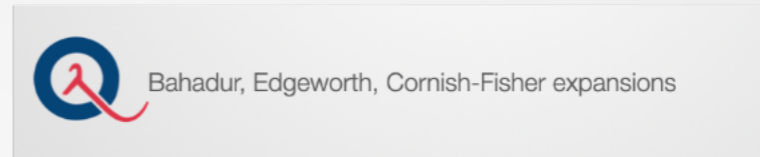


Presented with mind



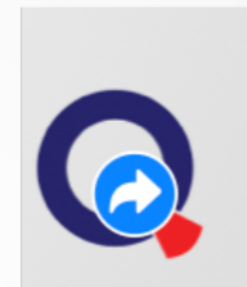
Data of Quantlet

Quantlets

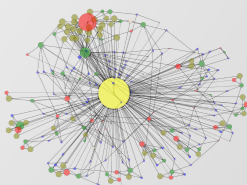
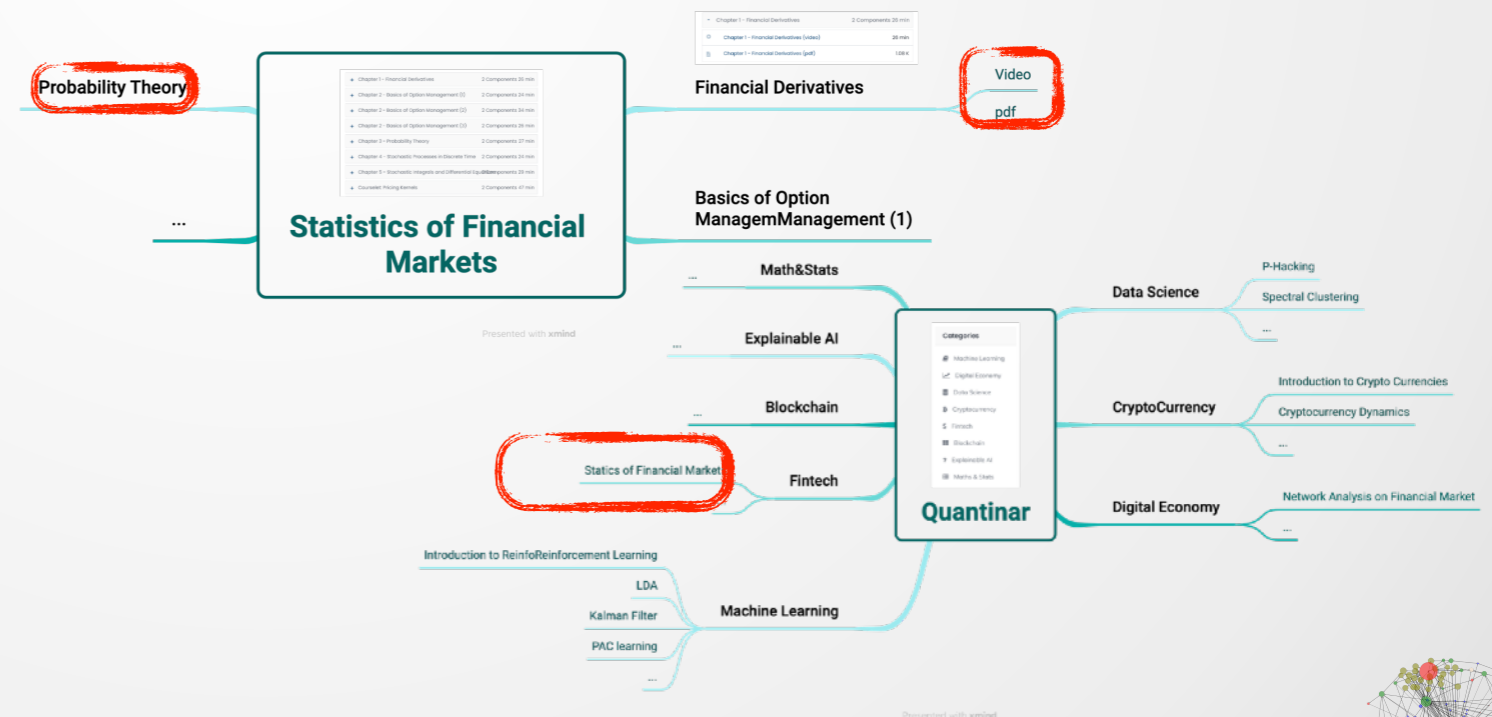


linkage to Quantinar: other courselet

pdf of courselet

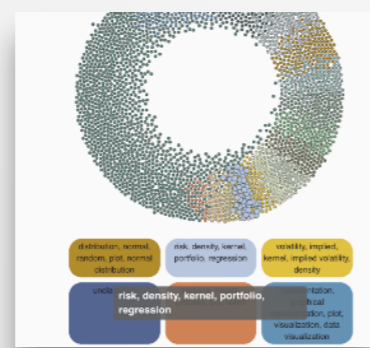
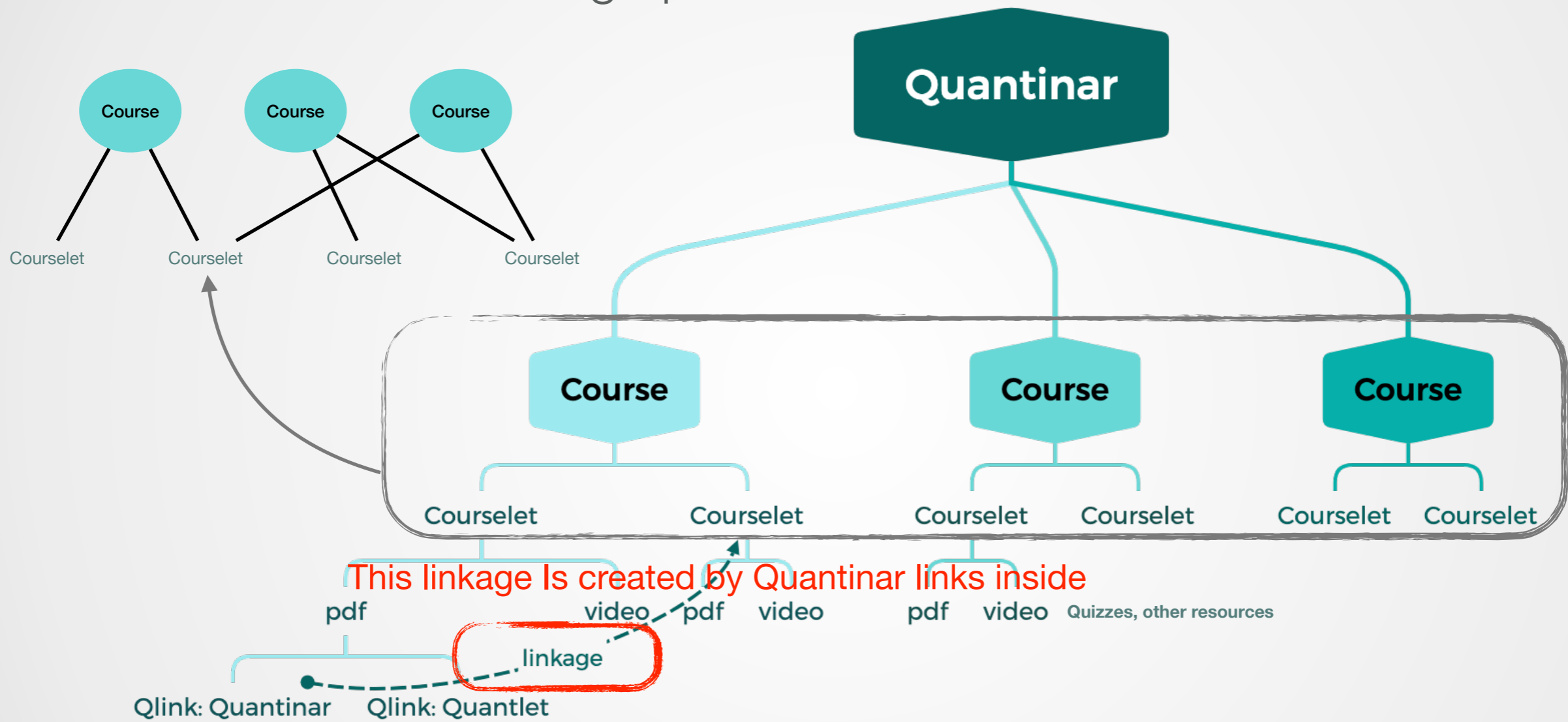


linkage to Quantlet: codes on github



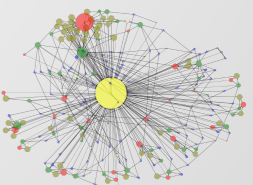
Graph Data Structure

- Then there comes a graph!



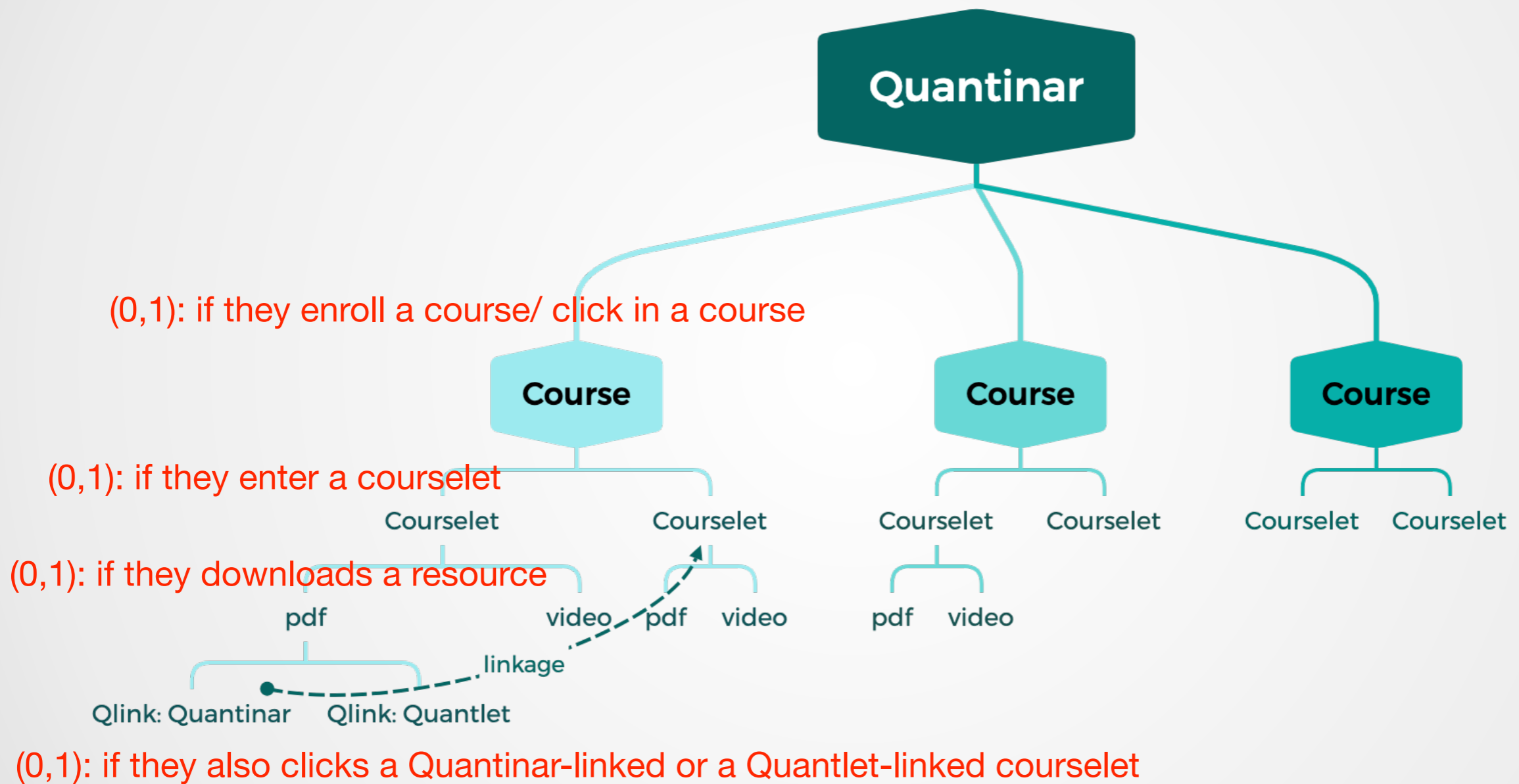
Presented with xmind

Quantlet also creates linkage

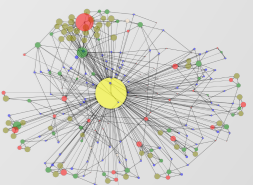


Student's Behavior

- 0: not click; 1: click

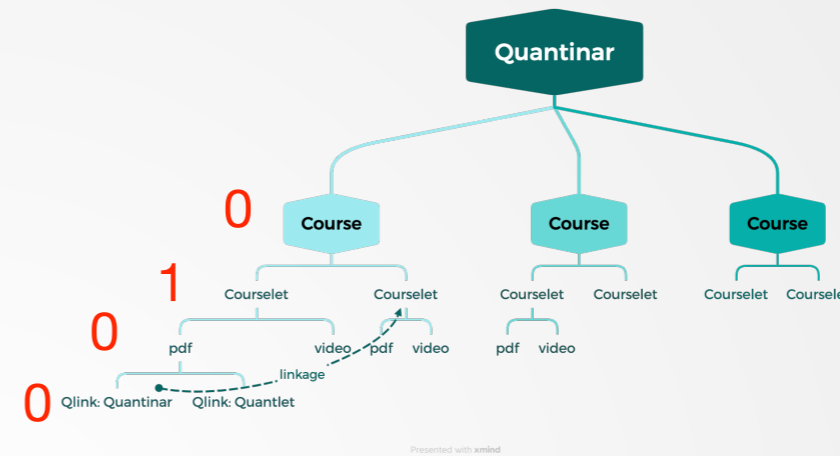


Presented with xmind



Constructing a Graph: OWL

- ▣ Ontologies: Courselets
- ▣ Relationship
 - ▶ Belong to one same course
 - ▶ Quantinar link related
 - ▶ Quantlet link related
- ▣ Properties: (0,1) - clicked or not; other info...



Property assertions: MaoKongLanChe

Object property assertions +

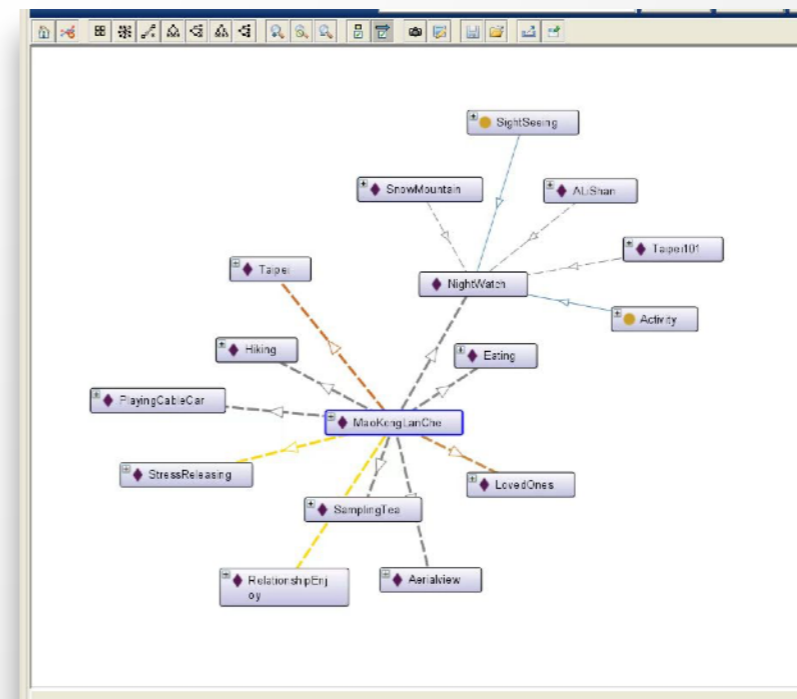
- hasActivity NightWatch
- bestFor LovedOnes
- hasActivity SamplingTea
- hasActivity Eating
- hasActivity Hiking
- hasPurpose StressReleasing
- hasActivity PlayingCableCar
- hasActivity Aerialview
- hasPurpose RelationshipEnjoy
- locatedIn Taipei

Data property assertions +

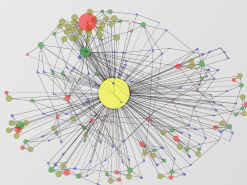
- hasYCoordinate 24.98261
- hasName "猫空茶园"^^string
- hasDescriptor "缆车"^^string
- requireTime "4"^^int
- hasStar "4"^^int
- hasXCoordinate 121.58409

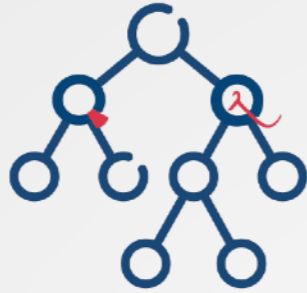
Negative object property assertions +

Negative data property assertions +

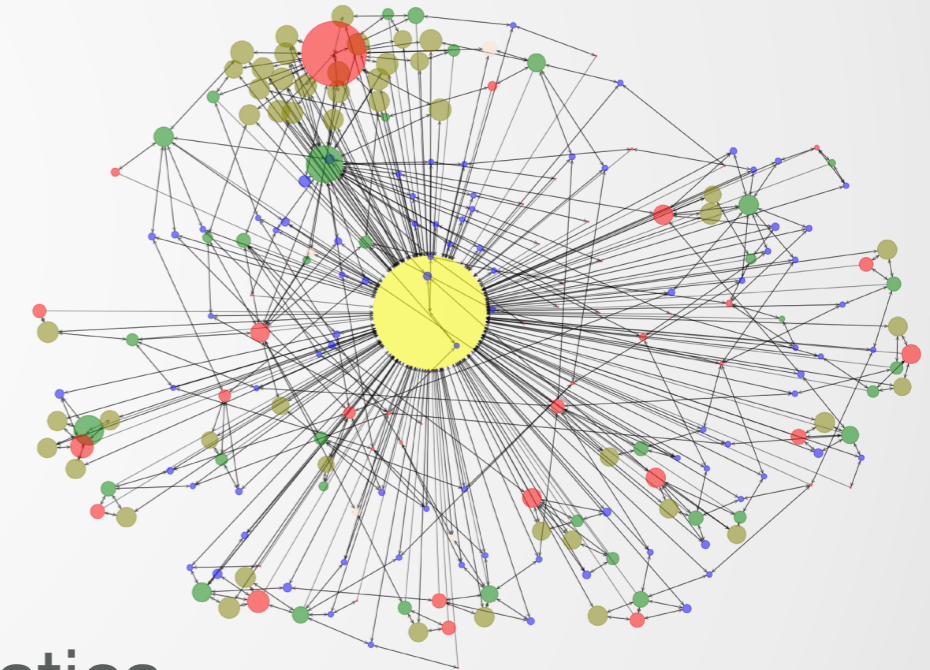


Those are owl constructions of our previous work on Taiwan Tourism Recommendation





Data. Analytics. Dissemination.



Wolfgang Karl Härdle
Xiaorui ZUO

Ladislaus von Bortkiewicz Professor of Statistics
Humboldt-Universität zu Berlin
Fudan University, Shanghai
theIDA.net

ASE, ACI NUS, NYCU 玉山學者