The Fundamental Principle of Data Science

Harry Crane

Department of Statistics
Rutgers

May 7, 2018

Web : www.harrycrane.com
Project : researchers.one
Contact : @HarryDCrane

Overview: “Foundations of Data Science”

Burning Questions

What is Data Science?

What is a Foundation?

What is the purpose of a Foundation?

Acknowledgments

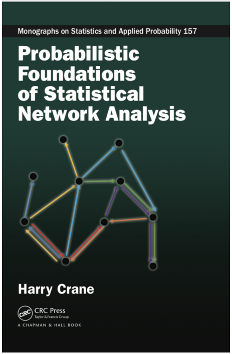
FPDS based on ongoing collaboration with Ryan Martin. Also very grateful for conversations with Glenn Shafer, Dimitris Tsementzis, and Snow Zhang.

Main References

- H. Crane. (2018). Logic of probability and conjecture.

  https://philpapers.org/rec/CRALOP

- H. Crane. (2018). *Probabilistic Foundations of Statistical Network Analysis*. Chapman–Hall.



Released May 2018

Foundations of Probability

$$\boxed{\text{Axioms/ Principles}} \xRightarrow{\text{encodes}} \boxed{\text{Formal Theory}} \xRightarrow{\text{expresses}} \boxed{\text{Basic Concept}} \xLeftarrow{\text{discerns}} \boxed{\text{Pre-formal Intuition}}$$

Foundations of Probability

Axioms/ Principles $\xRightarrow{\text{encodes}}$ Formal Theory $\xRightarrow{\text{expresses}}$ Basic Concept $\xLeftarrow{\text{discerns}}$ Pre-formal Intuition

"Frequentist Probability"

Kolmogorov's Axioms $\Longrightarrow$ Frequentist Probability $\Longrightarrow$ Frequencies $\Longleftarrow$ Chance/ Randomness

"Bayesian Probability"

de Finetti's coherence $\Longrightarrow$ Subjective Probability $\Longrightarrow$ Prices of fair bets $\Longleftarrow$ Credence

"Dempster–Shafer Probability"

Axioms of Belief Functions $\Longrightarrow$ D–Shafer Theory $\Longrightarrow$ Measure of evidence $\Longleftarrow$ Degrees of Belief

“Foundations” of Statistics

Axioms/ Principles $\xRightarrow{\text{encodes}}$ Formal Theory $\xRightarrow{\text{expresses}}$ Basic Concept $\xRightarrow{\text{supports}}$ Inference Methods

“Frequentist Statistics”

Kolmogorov’s Axioms $\Longrightarrow$ Frequentist Probability $\Longrightarrow$ Frequencies $\Longrightarrow$ P-values, Conf intervals, etc.

“Bayesian Statistics”

de Finetti’s coherence $\Longrightarrow$ Subjective Probability $\Longrightarrow$ Prices of fair bets $\Longrightarrow$ Bayes Factor, Posterior inf

“Inferential Models”

Axioms of Belief Functions $\Longrightarrow$ D–Shafer Theory $\Longrightarrow$ Measure of evidence $\Longrightarrow$ Valid prob. inference

Instances of Data Science

$$\boxed{\text{Data}} \xRightarrow{\text{Method}} \boxed{\text{Inference}} \xRightarrow{\text{Rule}} \boxed{\begin{array}{c}\text{Decision/}\\ \text{Application}\end{array}}$$

Instances of Data Science

Data $\xRightarrow{\text{Method}}$ Inference $\xRightarrow{\text{Rule}}$ Decision/ Application

Hypothesis testing

“Set” of Observations $\Longrightarrow$ P-value $\Longrightarrow$ “Reject” H_0

Instances of Data Science

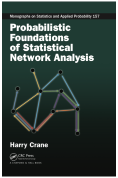
Data $\xRightarrow{\text{Method}}$ Inference $\xRightarrow{\text{Rule}}$ Decision/ Application

Hypothesis testing

"Set" of Observations $\Longrightarrow$ P-value $\Longrightarrow$ "Reject" H_0

Network science

Interaction data $\Longrightarrow$ Network analysis $\Longrightarrow$ Advertising campaign



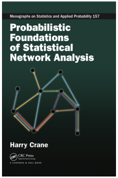
Data $\xRightarrow{\text{Method}}$ Inference $\xRightarrow{\text{Rule}}$ Decision/ Application

Hypothesis testing

"Set" of Observations $\Longrightarrow$ P-value $\Longrightarrow$ "Reject" H_0

Network science

Interaction data $\Longrightarrow$ Network analysis $\Longrightarrow$ Advertising campaign



Clustering/Classification

Cloud of Points $\Longrightarrow$ Clustering $\Longrightarrow$ Fraud Detection

The F-word

$$\textit{Data} \xrightarrow{\textit{Method}} \textit{Inference} \xrightarrow{\textit{Rule}} \textit{Decision}$$

- Methods — Hypothesis tests, Bayesian methods, confidence distributions, fiducial distributions, machine learning — are **not** Foundations.
- Foundations cannot only be based on statistics and probability.
  - Algorithmic methods (clustering, network science, machine learning) are data science.
  - Inferential Models based on Dempster–Shafer theory – not strictly probability but still statistics/data science.
- Foundations of Data Science should handle **all kinds of (complex) data** – not just sets, sequences, tables, arrays, etc.

The F-word

$$\textit{Data} \xrightarrow{\textit{Method}} \textit{Inference} \xrightarrow{\textit{Rule}} \textit{Decision}$$

- Methods — Hypothesis tests, Bayesian methods, confidence distributions, fiducial distributions, machine learning — are **not** Foundations.
- Foundations cannot only be based on statistics and probability.
  - Algorithmic methods (clustering, network science, machine learning) are data science.
  - Inferential Models based on Dempster–Shafer theory – not strictly probability but still statistics/data science.
- Foundations of Data Science should handle **all kinds of (complex) data** – not just sets, sequences, tables, arrays, etc.

Role of a Foundation:

- **Internal** (Logical): Provides a '**standard of coherence**' by which to assess methods.
  - Frequentist statistics: *ad hoc* principles/rules of thumb.
  - Bayesian statistics: coherence/Dutch book argument.
  - Inferential Models (Martin & Liu): 'validity' as standard of rigor.
- **External** (Scientific): Provides a '**standard of falsification**' for inferences.
  - Inferences are meaningful/scientific only if they are falsifiable.
  - Any method for data science should have built-in criteria under which inferences would be falsified.

$$\textit{Data} \xrightarrow{\textit{Method}} \textit{Inference} \xrightarrow{\textit{Rule}} \textit{Decision}$$

Basic Concepts of Data Science:

1. **Fundamental Principle of Data Science** (Crane–Martin): A **Method** is a protocol which takes any piece of data to an inference about every assertion A.

$$\mathbf{Method} : \mathbf{Data} \to \forall A\ \mathbf{Inference}(A).$$

2. Identical data structures should lead to identical inferences:

$$D =_{\mathbf{Data}} D' \quad \Longrightarrow \quad \forall A\ \mathbf{Method}_D(A) =_{\mathbf{Inference}(A)} \mathbf{Method}_{D'}(A).$$

How to make these principles rigorous?

$$\text{Data} \xrightarrow{\text{Method}} \text{Inference} \xrightarrow{\text{Rule}} \text{Decision}$$

If the data provides sufficient evidence in favor of A, then inferences should support A.

Hypothesis testing:

$$H_0 : \text{null hypothesis}$$

- Method: compute a P-value based on the data D:

$$P = \Pr(D \mid H_0).$$

- Interpret '$P < \alpha$' as evidence against H_0 (i.e., reject H_0).
- Inference based on P:

$$\textbf{Method}_D(H_0) = \begin{cases} \text{reject } H_0, & P < \alpha, \\ \text{inconclusive}, & \text{otherwise.} \end{cases}$$

Simple Example: Hypothesis testing (Bayes)

$$\text{Data} \xrightarrow{\text{Method}} \text{Inference} \xrightarrow{\text{Rule}} \text{Decision}$$

If the data provides sufficient evidence in favor of A, then inferences should support A.

Hypothesis testing:

$$H_0 : \text{null hypothesis}$$
$$H_1 : \text{alternative hypothesis}$$

- Method: compute a Bayes factor based on the data D:

$$BF = \frac{\Pr(D \mid H_1)}{\Pr(D \mid H_0)}.$$

- Interpret BF as measure of evidence in favor of H_1 over H_0 (denoted as A).
- Inference based on P (from Kass & Raftery, 1995):

$$\textbf{Method}_D(A) = \begin{cases} \text{inconclusive}, & BF < 3.2, \\ \text{substantial}, & 3.2 \leq BF \leq 10, \\ \text{strong}, & 10 < BF \leq 100, \\ \text{decisive}, & BF > 100. \end{cases}$$

Formalization: Data as Structure

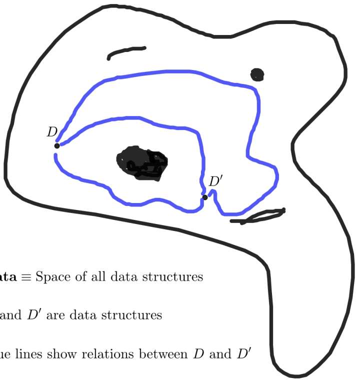
$$\textit{Data} \xrightarrow{\textit{Method}} \textit{Inference} \xrightarrow{\textit{Rule}} \textit{Decision}$$

- Informal: **Data** is **observable structure**, including measurements, relations between measurements (interactions, dependencies), etc.

Formalization: Data as Structure

$$\textit{Data} \xrightarrow{\textit{Method}} \textit{Inference} \xrightarrow{\textit{Rule}} \textit{Decision}$$

- Informal: **Data** is **observable structure**, including measurements, relations between measurements (interactions, dependencies), etc.
- Formal: Represent **Data** as a **topological space** whose points consist of observable information (observable 'data') and the relations between them.



Data $\equiv$ Space of all data structures

D and D' are data structures

Blue lines show relations between D and D'

Formalization: The structure of evidence

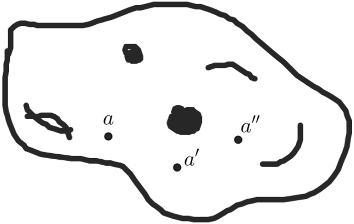
$$\textit{Data} \xrightarrow{\textit{Method}} \textit{Inference} \xrightarrow{\textit{Rule}} \textit{Decision}$$

- Informal: Inferences are based on the '**body of evidence**' in support of a claim.

Formalization: The structure of evidence

$$\textit{Data} \xrightarrow{\textit{Method}} \textit{Inference} \xrightarrow{\textit{Rule}} \textit{Decision}$$

- Informal: Inferences are based on the '**body of evidence**' in support of a claim.
- Formal: Represent the body of evidence for A as a **topological space Evid**(A) whose points are pieces of evidence in favor of A and paths are relations between pieces of evidence.



$\mathbf{Evid}(A) \equiv$ Body of evidence supporting A

a, a', a'' are distinct pieces of evidence for A

Formalization: The logical structure of inference

$$\textit{Data} \xrightarrow{\textit{Method}} \textit{Inference} \xrightarrow{\textit{Rule}} \textit{Decision}$$

- Informal: An **inference** is a **judgment** about **evidential strength** (based on data).

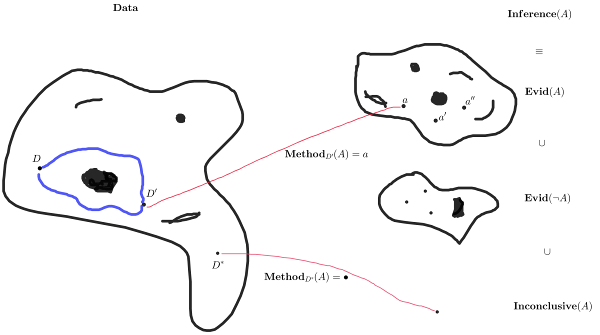
Formalization: The logical structure of inference

$$Data \xrightarrow{Method} Inference \xrightarrow{Rule} Decision$$

- Informal: An **inference** is a **judgment** about **evidential strength** (based on data).
- Formal: Represent **Inference**(A) as the **disjoint union of spaces** whose points are evidence in favor of A, in favor of $\neg A$, or that the outcome is inconclusive.

$$\textbf{Inference}(A) \equiv \textbf{Evid}(A) \cup \textbf{Evid}(\neg A) \cup \textbf{Inconclusive}(A).$$

- Can account for strength of evidence (strong, suggestive) but simple case here.



(1) For any assertion A there is a space of evidence for that assertion:

$$\frac{A \in \textbf{Space}}{\textbf{Evid}(A) \in \textbf{Space}}$$

(2) If A has been proven, then there is evidence to support A:

$$\frac{\begin{array}{c} A \in \textbf{Space} \\ a \in A \end{array}}{\texttt{evid}_A(a) \in \textbf{Evid}(A)}$$

(3) If A implies B, then evidence for A implies evidence for B:

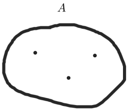
$$\frac{\begin{array}{c} A \in \textbf{Space} \qquad B \in \textbf{Space} \\ f : A \to B \end{array}}{\texttt{imp}(f) : \textbf{Evid}(A) \to \textbf{Evid}(B)}$$

(4) There is no evidence for a contradiction ($\emptyset$):

$$\frac{}{\textbf{Evid}(\emptyset) \equiv \emptyset}$$

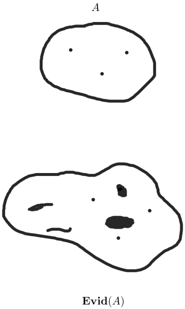
The Logic of Evidence: Syntax (Crane, 2018)

$$\frac{A \in \textbf{Space}}{} \qquad (1)$$

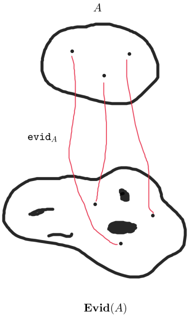


The Logic of Evidence: Syntax (Crane, 2018)

$$\frac{A \in \textbf{Space}}{\textbf{Evid}(A) \in \textbf{Space}} \tag{1}$$



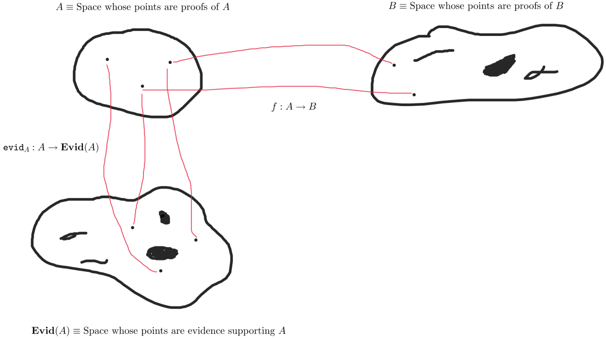
$$\frac{A \in \textbf{Space}}{\textbf{Evid}(A) \in \textbf{Space}} \quad (1) \qquad \frac{a \in A}{\texttt{evid}_A(a) \in \textbf{Evid}(A)} \quad (2)$$



The Logic of Evidence: Syntax (Crane, 2018)

$$\frac{A \in \textbf{Space}}{\textbf{Evid}(A) \in \textbf{Space}} \quad (1) \qquad \frac{a \in A}{\texttt{evid}_A(a) \in \textbf{Evid}(A)} \quad (2)$$

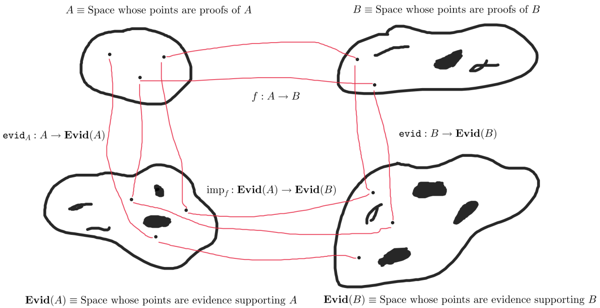
$$\frac{A \in \textbf{Space} \qquad B \in \textbf{Space} \qquad f : A \to B}{} \quad (3)$$



The Logic of Evidence: Syntax (Crane, 2018)

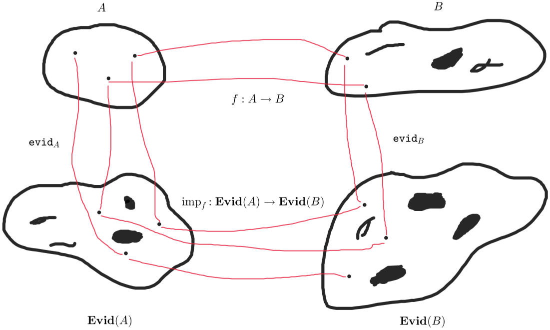
$$\frac{A \in \textbf{Space}}{\textbf{Evid}(A) \in \textbf{Space}} \quad (1) \qquad \frac{a \in A}{\texttt{evid}_A(a) \in \textbf{Evid}(A)} \quad (2)$$

$$\frac{A \in \textbf{Space} \qquad B \in \textbf{Space} \qquad f : A \to B}{\texttt{imp}(f) : \textbf{Evid}(A) \to \textbf{Evid}(B)} \quad (3)$$



$$\frac{A \in \textbf{Space}}{\textbf{Evid}(A) \in \textbf{Space}} \quad (1) \qquad\qquad \frac{a \in A}{\mathtt{evid}_A(a) \in \textbf{Evid}(A)} \quad (2)$$

$$\frac{A \in \textbf{Space} \qquad B \in \textbf{Space} \qquad f : A \to B}{\mathtt{imp}(f) : \textbf{Evid}(A) \to \textbf{Evid}(B)} \quad (3) \qquad\qquad \frac{}{\textbf{Evid}(\emptyset) \equiv \emptyset} \quad (4)$$



Consequences of these axioms

The following can be derived as **theorems** from the rules (1)-(4).

- $\mathbf{Evid}(A \cap B) \to \mathbf{Evid}(A) \cap \mathbf{Evid}(B) \to \mathbf{Evid}(A) \cup \mathbf{Evid}(B) \to \mathbf{Evid}(A \cup B)$

  Inferences supporting 'A and B' (jointly) support inferences for A and for B (individually), which in turn supports inferences for A or inferences for B, which supports inference for 'A or B'. (Arrows do not reverse in general.)

Universal/Existential quantification:

- $\mathbf{Evid}\,(\forall a \in A,\ B(a)) \to \forall a \in A,\ \mathbf{Evid}(B(a))$.

  Inference that $B(a)$ holds for all $a \in A$ supports inference that $B(a)$ holds for every $a \in A$.

- $\exists a \in A,\ \mathbf{Evid}(B(a)) \to \mathbf{Evid}\,(\exists a \in A,\ B(a))$.

  Inferring $B(a)$ for some $a \in A$ supports inference that there exists $a : A$ such that $B(a)$ holds.

Derived inference rules:

- $\mathbf{Evid}(A) \cap (A \to B) \to \mathbf{Evid}(A \cap B)$

  Inferring A and proving that A implies B supports inference in A and B.

- $A \cap \mathbf{Evid}(A \to B) \to \mathbf{Evid}(A \cap B)$.

Formalization in Coq/UniMath: Definition of **Evid** as a homotopy type

$$\frac{A \in \textbf{Space}}{\textbf{Evid}(A) \in \textbf{Space}} \quad (1)$$

$$\frac{\begin{array}{c} A \in \textbf{Space} \\ a \in A \end{array}}{\texttt{evid}_A(a) \in \textbf{Evid}(A)} \quad (2)$$

$$\frac{\begin{array}{c} A \in \textbf{Space} \qquad B \in \textbf{Space} \\ f : A \to B \end{array}}{\texttt{imp}(f) : \textbf{Evid}(A) \to \textbf{Evid}(B)} \quad (3)$$

$$\frac{}{\textbf{Evid}(\emptyset) \equiv \emptyset} \quad (4)$$

Written in Coq using UniMath library:

```
(** * Foundations of Data Science
Harry Crane
Started on May 1, 2018
Preparation for BFF 5 talk at University of Michigan on "The Fundamental Principle of Data Science"

Implementation of basic theorems from H. Crane.  (2018).  Logic of probability and conjecture.

https://philpapers.org/rec/CRALOP
*)


(** *** Settings *)

Unset Automatic Introduction.
(* The above line has to be removed for the file to compile with Coq8.2 *)

(** *** Imports *)

Require Export UniMath.Foundations.Preamble.
Require Export UniMath.Foundations.PartA.
Require Export UniMath.Foundations.UnivalenceAxiom.

(* end of "Preamble" *)

(* Define Evidence type former *)

Definition Space := UU.

Inductive Evid (A : Space) : Space :=
| evid : A → Evid A.

(* Elimination, Computation, and Zero rules for Evidence Type *)

Hypothesis  imp: ∏  (A : Space) (P : Evid A → Space),
    (∏ a : A, P (evid A a)) → ∏ p : Evid A, Evid (P p).

Hypothesis  comp : ∏  (A : Space) (P : Evid A → Space)
  (d : (∏ a : A, P (evid A a))), (∏ a:A, (imp A P d) (evid A a) = evid (P (evid A a)) (d a)).

Hypothesis evid_zero : Evid ø = ø.

(* Special case of imp for nondependent types *)

Definition  Prob_imp_nondep : ∏  {A B : Space}, (A → B) → (Evid A → Evid B).
Proof.
  intros A B f a.
  apply (imp A (λ _, B) (λ a, f a)) in a.
  apply a.
Defined.

∏U:**-  FPDS-BFF5-1    Top L11    (Coq Script(0-) Holes)
```

Formalization in Coq/UniMath: Proofs of basic properties

Theorems mentioned earlier:

$$\textbf{Evid}(A \cap B) \to \textbf{Evid}(A) \cap \textbf{Evid}(B)$$

$$\textbf{Evid}(A) \to \textbf{Evid}(A \cup B)$$

$$\textbf{Evid}(B) \to \textbf{Evid}(A \cup B)$$

$$\textbf{Evid}(A) \cup \textbf{Evid}(B) \to \textbf{Evid}(A \cup B)$$

Written in Coq using UniMath library:

```
(* Theorems for  product and coproduct *)

Theorem split_prob: ∏  (A B : Space), Evid ( A × B ) → Evid A × Evid B.

Proof. intros A B.  intros x.
      assert (H: Evid (A × B) → Evid A). { intros.
                                    apply (@Prob_imp_nondep (A × B) A pr1) in x. apply x. }

      assert (H': Evid (A × B) → Evid B). {intros.  apply (@Prob_imp_nondep (A × B) B pr2) in x. apply x. }
      split.
      - apply H in x. apply x.
      - apply H' in x. apply x.
Qed.

Notation "A ∐ B" := (coprod A B) (at level 20): type_scope.

Theorem inl_prob : ∏  {A B : Space}, Evid A → Evid ( A ∐ B ).

Proof.
  intros A B.  intros x.
  Check Prob_imp_nondep.
  apply (@Prob_imp_nondep A (A ∐ B) (@inl A B)) in x.
  apply x.
Qed.

Theorem inr_prob : ∏  {A B : Space}, Evid B → Evid ( A ∐ B ).

Proof.
  intros A B.  intros x.
  apply (Prob_imp_nondep (@inr A B)) in x.
  apply x.
Qed.

Theorem prod_co_prob : ∏  (A B : Space), Evid ( A × B ) → Evid ( A ∐ B ).

Proof. intros A B.  intros x.
      apply (@Prob_imp_nondep ( A × B ) (coprod A B) (fun y => @inl A B (pr1 y))) in x. apply x.
Qed.

Theorem join_prob : ∏  (A B : Space), Evid A ∐ Evid B → Evid ( A ∐ B ).
Proof.
  intros A B. intros x.
  destruct x.
  - apply inl_prob. apply e.
  - apply inr_prob. apply e.
Qed.
```

Interpretation of **Evid**: Semantics

Can these properties be interpreted in a familiar statistics/data science context?

Can interpret the space of evidence in terms of

- Probability functions on fixed set Ω.
- Spaces of probability measures (Giry monad in Category Theory).
- Similarly, finitely additive probabilities as in 1 and 2.
- Dempser–Shafer belief functions with fixed frame of discernment Ω.
- Spaces of Dempster–Shafer functions.

Immediate connections to

- Maximum likelihood estimation
- Hypothesis testing: P-values, Bayes factors
- Inferential Models

Next Steps

There are many ...

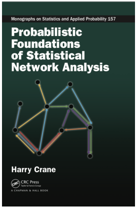
- Develop the theory further.
- Shore up the semantics in terms of existing approaches.
- Formalize existing statistics, machine learning, algorithmic approaches in this framework.
- Many more ....

References

- H. Crane. (2018). Logic of probability and conjecture.

  https://philpapers.org/rec/CRALOP
- H. Crane. (2017). Probability as Shape. Talk at BFF4. Video available at

  https://youtu.be/cnDKO0jY6dY
- H. Crane. (2018). *Probabilistic Foundations of Statistical Network Analysis*. Chapman–Hall.
- D. Tsementzis. (2018). A meaning explanation for HoTT. *Synthese*, to appear.



Web : www.harrycrane.com

Project : researchers.one

Contact : @HarryDCrane