idee per corsi secondo anno:

* **Explainable AI**
* **Ethics**
* **Law**
* **Neural Nwtroks (Ansuini)**
* **Data Visualization**

Academic English at the University of Trieste Katia Peruzzo 30

**“community detection” + “self supervised learning”**

Artificial Intelligence in society Arnaldi 1 10

Law and AI Infantino+guests 2,5 20 spring

Explainable AI Bortolussi + guests 1,5 12 may-ju

Data Visualization Tusar 2

**>>>>>>>>>> Neural / Deep con Alessio Ansuini <<<<<<<<<<<<< quanto tempo dedica a self supervised…**

Paper

[Statistical Methods & Applications | Home (springer.com)](https://www.springer.com/journal/10260)

[Statistical Methods & Applications | Editors (springer.com)](https://www.springer.com/journal/10260/editors)

[Get your manuscript ready for journal editors | Springer Nature](https://beta.springernature.com/pre-submission?journalId=10260)

* title
* abstract
* introduction
* methods
* results
* discussion

**Introduction**

Why we need to analyze this data Parto dai dati e dall’esigenza di analisi (univ vs altri)

Letteratura – what has been done before

**Data**

* the dataset is Laureati STEM
* The resulting network is fractioned in different components (a giant component and X small ocmponents with up to 4 nodes). The GC has very high density
* Weights can be set according to different strategies. Eg w=1, W=number of years, also non linear: W = number of years discounted as in NVP. W limited to max 3 years and min duration.
* Once we defined the weights, we can analyse Coreness and Stergth (start with a definition). Compare using NMI (and define NMI)
* We need a partition of the network: we use modularity-based algorothms (first define Modularity)

**Methodology**

* First: Louvain. A single iteration provides a result, but it is greedy, so it is a local max of modularity. A more informative approach is to repeat N times, which leads to N local max. We can then explore the distribution of modularity and either
  + Select the iteration with max modularity
  + Perform a consensus between the solutions with modularity above median

The former approach generates a partition of the network, where each node is assigned to a cluster). It is better than a single run, but it disregards the information in the distribution (FIGURE: distribution of modularity over N trials)

The latter approach, more complex and demanding computation, allows to calculate the probability of membership.

We adopted the latter approach, improved by inducing further variability as in “random forest”: in each trial a 5% of the edges has weight <- 0

Thus we can calculate the probability (let them vote, as in RF)

Pseudocodice della procedura che fa la gestione della variabilità

**Results**

Da pagina 9 in poi

**Conclusions**

*SOLO algoritmi con modularity e reti PESATE e UNDIRECTED*

* **Usiamo louvain**
* **Valorizziamo Leiden con parametri**
* **Ne consideriamo anche altre esplicitamente basate su modularity**
* **Non usiamo EV e EB**

**​​​​​​​**

**FRINGE** [(PDF) FRINGE: A New Approach to the Detection of Overlapping Communities in Graphs (researchgate.net)](https://www.researchgate.net/publication/221433189_FRINGE_A_New_Approach_to_the_Detection_of_Overlapping_Communities_in_Graphs)

**C’è troppa roba, è eterogeneo e forse 2 paper ci stanno**

* **Consensus clustering su UN SOLO METODO. C’è una variabilità di risultati su Louvain e applico un consensus clustering sulle 1000 iterazioni di una sola procedura.**
* **Partiamo da pag 9 risultati, questa è discussion of results.**
* **Spiegare meglio nel metodo la Comm Zero**

**Perchè fai tutto ciò?**

**Motivare bene 2 cose**

* **Perché RF dovrebbe funzionare? Devo frazionare reti molto dense. Quindi annullo alcuni legami a caso come in RF**
* **Posso prendere quello con la modularity più alta. Ma mi perderei alcune informazioni. E così mi posso calcolare la probability of assignation**
* **Perché prendo il confronto con la modularity più alta**
* **Non mi va bene la single run, perché non so se ha una buona modularity o meno**
* **E’ unsupervised learning**
* **Applichiamo il concetto di consensus**
* **Pseudocodice fino a pagina 9**
  + **Weigth**
  + **Resolution**
  + **3 parametri per community zero**

**PSEUDOCODICE da fare**

**Partiamo dai DATI e dal problema di interpretare le relazioni tra università.**

**Gestire la variabilità delle soluzioni è un problema NOTO tra i practitioners. La libreria igraph non ti permette di settare il seed . e se lo permettesse sarebbe anche peggio.**

**Tuning della resolution**

**Interpretazione dei risultati:**

**La 13 è bella**

**“i dati sono complessi, la rete è DENSA e il clustering non viene molto chiaro con EB e EV” quindi modularity based clstering,**

**metto al centro dei risultati le due università.**

* **Come si relazionano con altri / con SISSA /**

**Paper 1 paper 2 con template di PLOS verifico tempi e template**

**STRUTTURA Del paper da mettere anche nella relazione come abstract**

**Intro**

**Parto dai dati e dall’esigenza di analisi (univ vs altri)**

**Letteratura**

**Data**

**Laureati STEM**

**Componnet: la rete è frazionata e la GC è MOLTO DENSA**

**Discussione sui pesi**

**Definire Core Stergth Modularity e NMI**

**Methodology**

**Spiego: una sola iterazione non è otitmale e anche la max modularity è n local minima.   
Non quantifichi**

**Quidi random forest**

**Quindi probability**

**Pseudocodice della procedura che fa la gestione della variabilità**

**Results**

**Da pagina 9 in poi**

**NEXT papers**

* **I diversi pesi e il loro effetto sui risultati**
* **Diverse professioni**
* **Altri metodi di clustering es Louvain**

**L’immagine è per un PPT**