



UNIVERSITÀ DEGLI STUDI DI PALERMO

Dipartimento di Scienze Economiche, Aziendali e Statistiche

Master annuale di secondo livello in

Data Science and Big Data Analytics

Analisi di rete nell'abbandono studentesco tra il primo e il secondo anno: il caso dell'Università degli studi di Palermo

Tesi di:

Salvatore Marcantonio

Relatore:

Prof. Michele Tumminello

Tutor aziendale:

Dott.ssa Giuseppa Lenzo

Anno Accademico 2019/2020

A mio Papà

Nessuno ti chiamerà più Abbandonata

Is. 62,4

Ringraziamenti

Ringrazio di cuore,

Il coordinatore, la commissione esaminatrice e i docenti del Master per avermi dato la possibilità di frequentare ed apprendere e l'INPS per il sostegno economico;

I colleghi, in particolare Miriam e Roberto per il supporto morale;

I tutor accademico e aziendale, Michele e Giusi, per il supporto nella stesura della tesi;

I colleghi universitari Giuseppe, Roberto e Mario, per la fornitura dei dati di base;

Infine e soprattutto Barbara e Giosuè per aver supplito con l'affetto al tempo che le ho tolto loro.

Sommario

Ringraziamenti	4
Indice delle Tabelle	6
Indice delle Figure	8
INTRODUZIONE	9
I DATI	11
ANALISI DESCRITTIVA	13
La scelta dei CdS	21
ANALISI DELLE RETI	24
La rete bipartita.....	24
La proiezione sugli studenti	28
La rilevazione delle comunità sulla proiezione degli studenti	30
La comunità della proiezione degli studenti.....	32
La proiezione sugli insegnamenti	33
Elementi di machine learning.....	36
ANALISI DEI PERCORSI.....	37
L'analisi osservazionale: i percorsi del primo anno	39
La rete individuale.....	39
La rete individuale aggregata.....	40
Intervento: uno strumento del COT per l'orientamento in itinere.....	42
LE PROSPETTIVE	46
La prospettiva dello studente	46
La prospettiva del docente.....	46
La prospettiva del coordinatore e del gruppo AQ.....	46
La prospettiva del COT	47
CONCLUSIONI	48
Bibliografia.....	50

Indice delle Tabelle

Tab. 1 Tasso di abbandono nel triennio 2017/18-2019/201	13
Tab. 2 Distribuzioni assolute e percentuali dell'anno di immatricolazione e dell'abbandono condizionato all'anno di iscrizione	13
Tab. 3 Distribuzioni assolute e percentuali del genere e dell'abbandono condizionato al genere	13
Tab. 4 Distribuzioni assolute e percentuali del tipo di corso di immatricolazione e dell'abbandono condizionato al tipo di corso di immatricolazione	14
Tab. 5 Distribuzioni assolute e percentuali del tipo di diploma e dell'abbandono condizionato al tipo di diploma	14
Tab. 6 Distribuzioni assolute e percentuali del voto di diploma e dell'abbandono condizionato al voto di diploma	14
Tab. 7 Distribuzioni assolute e percentuali del numeri di esami superati al primo anno e dell'abbandono condizionato al numero di esami superati al primo anno	15
Tab. 8 Distribuzioni assolute e percentuali del numeri di esami superati al primo e dell'abbandono condizionato al numeri di esami superati al primo	15
Tab. 9 Distribuzione doppia assoluta del numero di esami superati al primo semestre e al secondo semestre	16
Tab. 10 Distribuzioni doppie assolute e percentuali del numero studenti condizionato al numero di esami superati al primo e al secondo semestre e dell'abbandono condizionato al numero di esami superati al primo e al secondo semestre	17
Tab. 11 Distribuzione assoluta, percentuale e percentuale cumulata del numero di studenti condizionato al numero di esami superati al primo semestre	18
Tab. 12 Distribuzione assoluta del numero di studenti condizionato al numero di esami contenenti i codici '6740' e '18683'	18
Tab. 13 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al numero di esami al primo semestre contenenti i codici '6740' e '18683'	18

Tab. 14 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al superamento di 2 esami.....	19
Tab. 15 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al superamento di 3 esami contenenti i codici '6740' e '18683'	19
Tab. 16 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al n° di esami contenenti i codici '6740' e '18683'	20
Tab. 17 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al n° di esami superati contenenti i codici '6740' e '18683'.....	20
Tab. 18 Eleggibilità dei CdS	21
Tab. 19 distribuzioni di alcune caratteristiche dei CdS selezionati	23
Tab. 20 Associazione tra comunità rilevate e abbandono.....	31
Tab. 21 Link statisticamente validati della rete sulla proiezione sugli insegnamenti.....	35
Tab. 22 Distribuzioni doppie assolute e percentuali del numeri studenti condizionato al superamento di 4 esami al variare del numero al primo e al secondo semestre e dell'abbandono condizionato al variare numero di esami superati al primo e al secondo semestre.....	43
Tab. 23 Motifs allargato a 3: suggerire ulteriore esame al 2° semestre.....	44
Tab. 24 Motifs allargato a 4: minimizzare la probabilità di abbandono.....	44

Indice delle Figure

Fig. 1 Rete bipartita.....	25
Fig. 2 Sottorete bipartita degli abbandoni.....	26
Fig. 3 Rete bipartita complementare	27
Fig. 4 Ego Rete bipartita complementare.....	28
Fig. 5 Rete proiezione sugli studenti della rete bipartita	29
Fig. 6 Distribuzione del grado dei nodi della proiezione studenti.....	30
Fig. 7 Frequenze assolute superamento degli esami condizionato distinte per comunità.....	32
Fig. 8 Rete proiezione sugli insegnamenti della rete bipartita.....	34
Fig. 9 Rete proiezione sugli insegnamenti statisticamente validata	35
Fig. 10 Rete astratta completamente connessa.....	38
Fig. 11 Esempio di percorso individuale osservato.....	40
Fig. 12 Esempio di percorso aggregato osservato.....	41
Fig. 13 Esempio di percorso previsionale aggregato	42
Fig. 14 Esempio di intervento finalizzato al percorso del secondo semestre.....	44
Fig. 15 Rete retrospettiva.....	45

INTRODUZIONE

Obiettivo di questa tesi è **l'analisi del fenomeno dell'abbandono** universitario nel passaggio tra il primo e il secondo anno attraverso l'applicazione dell'analisi di rete e la **definizione di azioni** incentrate nel percorso del secondo semestre atte a diminuirne l'intensità.

In questo contesto, **l'abbandono è definito** come la mancata iscrizione allo stesso Corso di Studio (CdS) di immatricolazione nell'anno accademico successivo all'immatricolazione. L'abbandono così definito comprende, oltre alla già accennata mancata iscrizione, anche il passaggio ad altro CdS dell'Ateneo differente da quello di immatricolazione oppure la formale rinuncia agli studi e immatricolazione ad altro CdS dell'Ateneo o nazionale nell'anno accademico successivo.

Tale definizione indirizza **l'analisi sul singolo CdS** poiché un passaggio o una rinuncia e nuova immatricolazione non rappresentano una perdita per l'Ateneo ma, d'altronde, il luogo naturale dove l'abbandono avviene è il CdS e a quel livello si deve informare e agire.

L'instabilità temporale dell'offerta formativa del primo anno, ad esempio cambio di semestre di un insegnamento da un anno all'altro, l'introduzione o sostituzione di insegnamenti, può porre un limite al potere predittivo dei modelli. L'argomento sarà precisato nel seguito.

Informalmente **una rete può essere definita come** un insieme di 3 elementi: un insieme di nodi, un insieme di archi o frecce congiungenti coppie di nodi e un layout che ne massimizzi la chiarezza espositiva (3). Più intuitivamente, una rete è un oggetto grafico formato tipicamente da palline connesse da linee, le proprietà visive delle palline e delle linee riflettono proprietà degli oggetti o delle relazioni tra oggetti.

Le reti sono attualmente utilizzate nell'analisi di vari fenomeni: sociali (reti sociali come facebook o twitter), di comunicazione (reti di comunicazione come il web o wikipedia), trasporti (reti di trasporto come i traffici aerei o metropolitani) o, infine, biologici (reti biologiche come l'interazione delle proteine in un determinato organismo vivente) (4).

Si consento una breve accenno sul perché della **scelta dello strumento**: certamente non mancano tecniche statistiche per la modellazione degli abbandoni: dalle semplici tabelle di frequenza condizionate, ai modelli ad equazioni come la regressione logistica, fino a ad altri modelli grafici come le bayesian networks. Caratteristica comune di tutte queste tecniche è la modellazione di relazioni tra variabili, ad esempio nella forma $Y = f(X_1, X_2, \dots, X_n)$, e la necessaria perdita dell'informazione (per aggregazione) a livello individuale. O anche l'implicita assunzione di indipendenza tra gli oggetti osservati (indipendenza per riga e modellazione della dipendenza per colonna). In una ideale matrice dei dati, l'analisi è condotta per "colonne".

Diversa è la prospettiva dell'analisi di rete, in cui si modella primariamente **la relazione individuale tra coppie oggetti** (il link) e a partire da questa si definiscono le relazioni macroscopiche (ad esempio la distribuzione del numero di link tra oggetti, ovvero il grado) o quelle mesoscopiche (ad esempio le comunità con opportune distinzioni tra i link interni alla comunità e quelli esterni ad essa).

In questo tipo di analisi le tradizionali variabili, ossia le caratteristiche degli oggetti, sono utilizzate come attributi dei nodi o delle relazioni in un ottica assortativa, in cui oggetti con caratteristiche simili tendono a sviluppare relazioni tra loro (omofilia) o viceversa, oggetti in relazione tra loro tendono uniformare le caratteristiche (influenza sociale).

Questo **cambio di prospettiva**, complementare a quelle standard, è stato l'evento che mi ha convinto ad indirizzare la tesi su questa tecnica.

D'altra parte però, è bene essere coscienti anche dei **limiti di tale strumento**. Almeno nelle applicazioni più diffuse, il primo limite è la restrizione della cardinalità delle relazioni a quella binaria, che può dare origine, ad esempio, al confondimento tra una originaria relazione ternaria e 3 relazioni binarie formalizzate nella rete, ovvero una generica relazione tra A, B e C si modella (e si confonde) con le tre relazioni binarie: A in relazione con B, B in relazione con C, C in relazione con A. In pratica, diverso è dire, e formalizzare, che A, B e C si sono incontrati simultaneamente dal dire, e formalizzare, che A si è incontrato con B, B con C, C con A, i quali potrebbero, o non potrebbero, essersi incontrati simultaneamente.. Comunque ai nostri scopi il limite è accettabile.

Infine è bene chiarire quale sia, oltre allo scopo e allo strumento, la **popolazione obiettivo**: immatricolati puri alle lauree triennali e magistrali a ciclo unico e iscritti per la prima volta a lauree magistrali secondo la definizione dell'anagrafe nazionale degli studenti universitari, soggetti a restrizioni di carriera come precisato nel seguito.

I DATI

I dati di base su cui si fonda l'analisi provengono da **due distinte fonti**: l'archivio **dell'offerta formativa di ateneo** (offweb.unipa.it) contiene gli insegnamenti del primo anno dei CdS dell'Ateneo e l'archivio delle **carriere degli studenti** (estratti dall'anagrafe nazionale degli studenti, ANS) contiene le informazioni sugli immatricolati, compreso il CdS di immatricolazione, e sugli esami superati durante il primo anno di immatricolazione.

Da queste fonti sono stati estratti le seguenti informazioni.

Dal database dell'**offerta formativa**:

- anno accademico erogazione;
- cds: codice, descrizione e classe;
- insegnamento: codice, descrizione, cfu, ssd, tipo, semestre erogazione.

Dopo una prima analisi su tutti i CdS, ne sono stati estratti 3 su cui costruire le reti e i percorsi analizzati.

Dal database **studenti**:

- identificativo anonimo dello studente;
- anno accademico immatricolazione;
- genere;
- codice cds di immatricolazione;
- diploma, tipo e voto (anche per classi) ;
- numero esami conseguiti nei due semestri e in totale;
- elenco esami conseguiti nei due semestri e in totale;
- abbandono o prosecuzione al secondo anno.

Si è deciso di includere i cosiddetti immatricolati puri o iscritti per la prima volta ad una LM¹ e conseguentemente escludere le seconde carriere gli studenti che abbiano acquisito almeno un esame convalidato o dispensato poiché al netto di convalide e dispense l'associazione tra performance al primo anno e abbandono o prosecuzione è più stretta.

Dal database **esami**:

- anno accademico esame;
- identificativo anonimo studente;
- codice cds di immatricolazione dello studente;
- esame: codice, cfu, semestre e sessione di assolvimento, voto, tipo (di base, caratterizzante,...).

Conformemente alla precedente esclusione, sono stati esclusi tutti gli esami convalidati o dispensati.

E' importante dichiarare subito che l'unità di misura del percorso dello studente utilizzata nel seguito sarà **il numero di esami superati e non**, come anche legalmente definito, **il numero di CFU acquisiti**. La scelta ha due ragioni: la prima vuole evidenziare come il superare, o non superare, un

¹ La definizione è ministeriale: gli studenti che per la prima volta si iscrivono ad un corso di studio universitario.

esame **ha un valore in se** non, o poco, influenzato dai CFU associati (un insegnamento da 9 CFU in un insieme di insegnamenti da 6 CFU non lo rende automaticamente una materia scoglio), la seconda ritiene il n° di esami **una misura più robusta** del n° di cfu poiché tipicamente un esame più equivalere da un minimo di 3 a un massimo di 12 CFU e la distribuzione della pesatura degli insegnamenti sia diversa da CdS a CdS.

ANALISI DESCRITTIVA

Prima di modellare l'abbandono con le reti, mostriamo alcune statistiche di base, calcolate su tutti gli immatricolati del triennio 2017/18 al 2019/20.

Innanzitutto la percentuale di abbandono nel triennio su tutti gli immatricolati

Il numero di immatricolati totali è di 28251, la **percentuale di abbandono si attesta al 24.1%**. Il dato è coerente con il valore riportato nella ultima scheda di monitoraggio annuale (SMA), di circa il 21%²

Tab. 1 Tasso di abbandono nel triennio 2017/18-2019/201

Abbandono	N°	%
Si	21378	75.7
No	6873	24.3

Le successive tabelle mostrano al centro la variabile di classificazione, a sinistra la numerosità e la percentuale rispetto al totale degli immatricolati e a destra la numerosità e la percentuale di abbandono condizionatamente alla variabile di classificazione.

Si nota una **piccola variabilità temporale** ma non così significativa da dovere trattare gli anni separatamente.

Tab. 2 Distribuzioni assolute e percentuali dell'anno di immatricolazione e dell'abbandono condizionato all'anno di iscrizione

N° Anno	% Anno	Anno	N° abbandoni	% abbandoni
7613	26.9	2017/18	1989	26.1
9889	35.0	2018/19	2384	24.1
10749	38.0	2019/20	2500	23.3

La distribuzione per genere mostra come abbandonano più le femmine, ma perché sono più numerose. **In percentuale abbandonano più i maschi**, la differenza supera il 6%.

Tab. 3 Distribuzioni assolute e percentuali del genere e dell'abbandono condizionato al genere

N° Genere	% Genere	Genere	N° abbandoni	% abbandoni
11765	41.6	M	3323	28.2
16486	58.4	F	3550	21.5

L'abbandono è molto più frequente, sia in valore assoluto che percentuale, tra gli immatricolati dei CdS post diploma, **lauree in primis e LMCU**, e meno tra gli immatricolati magistrali. Questi ultimi infatti hanno già una carriera universitaria triennale alle spalle e hanno subito una ulteriore selezione, infatti non tutti i laureati triennali proseguono alla magistrale.

Si noti come la percentuale di abbandono tra gli immatricolati delle lauree sfiori il 30%, mentre quelle alle LMCU si abbassano di 10 punti percentuali e di ulteriori 10 quella alle LM.

² L'indicatore è leggermente diverso, poiché definito in termini di Classe ed è calcolato solo sulle Lauree e Lauree magistrali a ciclo unico: percentuale di studenti che proseguono nel II anno nella stessa classe di laurea (IA14)

Un valore assoluto la numerosità degli abbandoni alla triennale è nettamente il dato da monitorare.

Tab. 4 Distribuzioni assolute e percentuali del tipo di corso di immatricolazione e dell'abbandono condizionato al tipo di corso di immatricolazione

N° Tipo CdS	% Tipo CdS	Tipo CdS	N° abbandoni	% abbandoni
18047	63.9	Laurea	5417	30.0
4302	15.2	Laurea Magistrale CU	889	20.7
5902	20.9	Laurea Magistrale	567	9.6

La distribuzione del tipo di diploma è fortemente dipendente dall'ambito del Cds di immatricolazione (per semplificare umanistico, scientifico, medico, sociale). Ciò nonostante anche a livello di ateneo si possono raccogliere alcune evidenze. I Licei classico e scientifico coprono il 50% degli immatricolati, e sono quelli in cui in proporzione si abbandona di meno, circa il 20%, mentre negli **istituti tecnici** la percentuale aumenta al 30, ed oltre negli **istituti professionali**, sebbene gli immatricolati siano numericamente al di sotto del 5%.

Tab. 5 Distribuzioni assolute e percentuali del tipo di diploma e dell'abbandono condizionato al tipo di diploma

N° Tipo DIPLOMA	% Tipo DIPLOMA	Tipo DIPLOMA	N° abbandoni	% abbandoni
5687	19,5	LICEI CLASSICI	1186	20,9
9508	32,6	LICEI SCIENTIFICI	1837	19,3
6279	21,4	ALTRI LICEI	1556	24,8
6080	20,9	ISTITUTI TECNICI	1903	31,3
1237	4,1	ISTITUTI PROFESSIONALI	445	36,0
413	1,4	ALTRO	114	27,6

I voti di diploma tendono leggermente verso le regioni basse, gli abbandoni sono **monotonamente discendenti all'aumentare del voto di diploma** (in classi).

Il range di percentuale di abbandono, compreso tra il 20% e il 30%, non è dissimile da quello generato dalle variabili di classificazione, ed induce a supporre che il voto di diploma abbia un limitato effetto predittivo sull'abbandono.

Tab. 6 Distribuzioni assolute e percentuali del voto di diploma e dell'abbandono condizionato al voto di diploma

N° Voto DIPLOMA	% Voto DIPLOMA	Voto DIPLOMA in Classi	N° abbandoni	% abbandoni
7590	26.9	[60-70]	2356	31.0
7987	28.3	(70-80]	2023	25.3
6437	22.8	(80-90]	1354	21.0
6237	22.1	(90-100]	1140	18.3

Il numero di esami superati, come è lecito attendersi, è di gran lunga **la variabile maggiormente predittiva dell'abbandono**.

Si noti innanzitutto come la moda della distribuzione del numero di esami sia 0, col 16.4%, e generi quasi un 90% di abbandoni. Inoltre il numero di abbandoni a 0 esami copre il 60% del totale degli abbandoni.

La percentuale di abbandono è monotonamente decrescente con il numero di esami superati. Tre esami è la prima frequenza, 14.4%, più bassa del valore medio di 24.3%, un ulteriore esame in più abbassa la percentuale a un valore molto più accettabile del 6.7%.

Il numero di esami al primo anno varia da CdS a CdS, la soglia di 8, in linea teorica 4 al primo semestre e 4 al secondo, è confortata anche dalla numerosità, l'ultima classe 8 o più esami ha frequenza simile a 1 o 2 esami.

Si noti come il potere discriminante di tutte le precedenti caratteristiche vari tra **un minimo del 20% a un massimo del 30%** di abbandoni, mentre quello del n° di esami **dal 11% al 89%**, segno che è il primo anno di carriera a fare la differenza, rispetto alla carriera precedente o alle caratteristiche personali,

Tab. 7 Distribuzioni assolute e percentuali del numero di esami superati al primo anno e dell'abbandono condizionato al numero di esami superati al primo anno

N° Studenti	% Studenti	N° esami al primo anno	N° abbandoni	% abbandoni
4732	16.47	0	4210	89.0
1887	6.7	1	1091	57.8
2118	7.5	2	656	31.0
2776	9.8	3	400	14.4
3408	12.1	4	229	6.7
3514	12.4	5	140	4.0
3807	13.5	6	72	1.9
3728	13.2	7	48	1.3
2281	8.1	8+	27	1.2

Accertato il valore predittivo del N° di esami, si è cercato di fare un passo in direzione dell'operatività mimando la tessa tabella ma con il n° di esami al primo semestre.

L'operatività va intesa in questo senso: la precedente tabella può essere calcolata solo a fine primo anno quando l'iscrizione al secondo è imminente o già fatta (ricordo che nei dati è compresa la sessione di settembre e che le iscrizioni iniziano da agosto per chiudersi a tipicamente dicembre).

La seguente tabella può essere calcolata già a fine primo semestre e dunque si può intervenire in itinere per migliorare la prestazione del secondo semestre e, se possibile, evitare l'abbandono.

La distribuzione del numero di esami è pressoché uniforme mentre la metà distribuzione degli abbandoni è sufficientemente discriminata. **Tre quarti di coloro che non superano alcun esame al primo semestre, indipendentemente da cosa farà al secondo, abbandona.** La percentuale si attesta al valore medio generale con un esame e si abbassa al 10% con due.

Tab. 8 Distribuzioni assolute e percentuali del numeri di esami superati al primo semestre e dell'abbandono condizionato al numeri di esami superati al primo semestre

N° Studenti	% Studenti	N° esami al primo semestre	N° abbandoni	% abbandoni
6151	21,8	0	4481	72,8
4862	17,7	1	1281	26,3
6409	22,7	2	660	10,3

N° Studenti	% Studenti	N° esami al primo semestre	N° abbandoni	% abbandoni
6122	21,7	3	325	5,3
4707	16,7	4+	126	2,7

Continuando su questa linea si è deciso di quantificare il valore predittivo del primo semestre rispetto al secondo con una tabella a doppia entrata. I risultati sono chiari:

Chi non ha dato nessun esame al primo semestre difficilmente non replica anche al secondo, sebbene pochi riescono anche ad arrivare a 4 o più.

Tra chi ha superato un esame al primo semestre, i più non hanno confermato la stessa performance. Circa il 50% arriva alla soglia dei 3 (1+>=2)

Chi ne ha superati 2 o 3 al primo semestre, conferma il numero anzi tende ad aumentare nel secondo. La diminuzione dei quelli che ne hanno superato 4 o più al primo semestre può essere solo apparente e dovuta alla strutturazione degli insegnamenti nei due semestri, ad esempio se il piano di studi prevede 4 insegnamenti al primo semestre e 3 al secondo necessariamente, anche superarli tutti implica una diminuzione al secondo semestre.

Tab. 9 Distribuzione doppia assoluta del numero di esami superati al primo semestre e al secondo semestre

		N° esami superati nel II semestre				
		0	1	2	3	4+
N° esami superati nel I semestre	0	4790	542	423	294	221
	1	1404	997	1141	877	661
	2	779	1153	1713	1440	1603
	3	334	721	1347	1813	2116
	4+	159	524	1147	1871	1134

Interessante valutare la frequenza di abbandono al variare del numero di esami sia la primo che al secondo semestre nonché la differenza di abbandono a parità di totale esami ma di diversa composizione semestrale.

Tra coloro che hanno superato un solo esame, sono più frequenti quelli che lo hanno superato al primo semestre (936 vs 181) e in percentuale abbandonano di più (66,7% vs 33,4%).

Il risultato potrebbe essere interpretato in questi termini: lo studente ha probabilmente ha seguito delle lezioni al primo semestre, è riuscito a superare un esame ma o ne ha dedotto che di quel passo non sarebbe riuscito a proseguire oltre ed ha abbandonato subito o pur frequentato delle lezioni del secondo non è riuscito ne a recuperare un esame del primo semestre ne a bissare la prestazione del primo ed ha abbandonato. Invece l'interpretazione di coloro che hanno superato un solo esame ma al secondo semestre potrebbe essere un serio impatto negativo con la didattica universitaria (diversa da quella secondaria superiore) al primo semestre, minimamente migliorato al secondo ma quanto basta per darsi una seconda possibilità al secondo anno.

Tra coloro che hanno superato due esami:

- entrambi al primo semestre, i più frequenti, abbandonano al 50%
- entrambi al secondo semestre, i meno frequenti, abbandonano al 17%
- uno al primo e uno al secondo, si attestano al valore medio del 25%

Anche tra coloro che hanno superato tre esami si consolida la tendenza appena vista, **maggiore è il numero degli esami superato al secondo semestre (a parità di totale nell'anno) minore è la**

frequenza di abbandono. E' pur vero che però la frequenza di coloro che superano tutti gli esami in un solo semestre è più bassa (circa 300) di color che superano in entrambi (circa 1100).

Tra coloro che hanno superato 4 o più esami, la frequenza di abbandono è bassa, ad eccezione di coloro che o hanno superato tali esami solo al primo semestre (30% di abbandono) o 3 al primo e uno al secondo (10%) ma sono numericamente pochi, in tutto il 3%. Potrebbe trattarsi anche di trasferimenti in uscita.

Tab. 10 Distribuzioni doppie assolute e percentuali del numero studenti condizionato al numero di esami superati al primo e al secondo semestre e dell'abbandono condizionato al numero di esami superati al primo e al secondo semestre

N° studenti	%° studenti	N° esami I semestre	N° esami II semestre	N° abbandoni	% abbandoni
4790	16,4	0	0	4246	88,6
542	1,9		1	181	33,4
423	1,4		2	73	17,3
294	1,0		3	24	8,2
221	0,8		4+	8	3,6
1404	4,8	1	0	936	66,7
997	3,4		1	245	24,6
1141	3,9		2	99	8,7
877	3,0		3	34	3,9
661	2,3		4+	8	1,2
779	2,7	2	0	364	46,7
1153	3,9		1	184	16
1713	5,9		2	101	5,9
1440	4,9		3	35	2,4
1603	5,5		4+	19	1,2
334	1,1	3	0	124	37,1
721	2,5		1	79	11
1347	4,6		2	76	5,6
1813	6,2		3	33	1,8
2116	7,2		4+	38	1,8
159	0,5	4+	0	47	29,6
524	1,8		1	25	4,8
1147	3,9		2	28	2,4
1871	6,4		3	26	1,4
1134	3,9		4+	8	0,7

Appare chiaro che raggiunta la soglia di quattro esami il rischio di abbandono si attesta a livelli fisiologici mentre al di sotto di tale soglia diventa discriminate quali esami sono stati superati poiché alcuni tendono ad associarsi maggiormente all'abbandono, altri meno.

A tal riguardo si analizzi il seguente caso:

Lo studente è iscritto a Giurisprudenza, ha sostenuto 2 esami, 6740 Storia del Diritto Romano e 18683 Inglese Giuridico.

Si inizia a posizionare il percorso rispetto alla propria coorte di immatricolati. Il 50% dei 560 immatricolati ha sostenuto al massimo due esami.

Tab. 11 Distribuzione assoluta, percentuale e percentuale cumulata del numero di studenti condizionato al numero di esami superati al primo semestre

N° esami I semestre	N° studenti	%	$\Sigma\%$
0	131	23,4	23,4
1	73	13,0	36,4
2	85	15,2	51,6
3	84	15,0	66,6
4	59	10,5	77,1
5	57	10,2	87,3
6	40	7,1	94,5
7	25	4,5	98,9
8	6	1,1	100,0
	560	100,0	

Si posiziona anche la coppia di esami sostenuti all'interno di insiemi più ampi. Sembra che la coppia sia più frequente in insiemi più ampi di esami che non sa soli.

Tab. 12 Distribuzione assoluta del numero di studenti condizionato al numero di esami contenenti i codici '6740' e '18683'

N° esami I semestre	N° studenti
2	12
3	23
4	20
5	19
6	19
7	11
8	6

Si prosegue valutando il percorso rispetto alle coorti precedenti.

Tab. 13 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al numero di esami al primo semestre contenenti i codici '6740' e '18683'

N° esami I semestre	N° studenti	% abbandono
2	177	6.8%
3	35	2.9%
4	17	0.0
5	19	5.3
6	15	6.7
7	10	0.0
8	14	14.3
9	1	0.0

Dopo di che si valuta l'aspetto qualitativo della coppia di esami.

Tab. 14 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al superamento di 2 esami

motif	N° studenti	% abbandono
18683_20633	1	100%
20633_20634	16	62.5
20633_20636	1	0
4035_6740	47	0
6740_10137	1	0
6740_15550	5	0
6740_18682	10	0
6740_18683	177	6.8
6740_20633	37	16.2
6740_20635	1	0
6740_20636	1	0
6740_2432	16	0

La coppia di esami è la più frequente in L'insegnamento 6740 si associa a basse la probabilità di abbandono, indipendentemente da con chi è in coppia la probabilità di abbandono è bassa

Gli insegnamenti 20633 o 20634 dovrebbero essere facili poiché anche se superati entrambi non proteggono dall'abbandono. Inoltre quando 6740 si associa a 20633 invece che a 18683 la frequenza di abbandono aumenta. Nulla si può dire sull'insegnamento 18683 poiché un'ulteriore associazione ha frequenza unitaria.

Inoltre si valuta l'allargamento del motif ad un ulteriore esame dello stesso semestre.

Tab. 15 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al superamento di 3 esami contenenti i codici '6740' e '18683'

motif	N° studenti	% abbandono
4035_6740_18683	16	0.0
6740_18683_10137	8	0.0
6740_18683_20633	8	12.5
6740_18683_2432	2	0.0
6740_18683_20635	1	0.0

L'allargamento produce ottimi risultati, solo un abbandono. L'ulteriore insegnamento più frequente è il 4035 con una percentuale di abbandono nulla.

Infine si valuta l'allargamento del motif a esami anche del II semestre.

Tab. 16 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al n° di esami contenenti i codici '6740' e '18683'

N° esami	motif	N° studenti	% abbandono
5	2432_4035_6740_15550_18683	64	0
6	2432_4035_6740_15550_18683_10137	7	0
8	2432_4035_6740_15550_18683_20633_20634_20635	1	0
6	2432_4035_6740_15550_18683_20635	1	0
4	2432_4035_6740_18683	1	100
4	2432_6740_15550_18683	38	0
5	2432_6740_15550_18683_10137	1	0
3	2432_6740_18683	4	0
4	4035_6740_15550_18683	19	5.3
5	4035_6740_15550_18683_10137	2	0
3	4035_6740_18683	2	0
3	6740_15550_18683	24	12.5
4	6740_15550_18683_10137	1	0
2	6740_18683	11	63.6
5	6740_18683_20633_20634_20635	1	0

Dei 177 casi del primo semestre, solo 11 non aggiungono altri esami e di questi 7 abbandonano. Il motif tende ad allargarsi da 4 a 6 esami a fine anno, con basse frequenze di abbandono.

In sintesi, si valuta la prosecuzione per ogni valore del numero di esami e si seleziona la frequenza massima di abbandono.

Tab. 17 Distribuzione assoluta del n° degli studenti e percentuale dell'abbandono condizionato al n° di esami superati contenenti i codici '6740' e '18683'

N° esami	MOTIFS	N° studenti	% abbandono
3	6740_18683_20633	1	100
6	6740_18683_20633_20634_20635_20636	2	50
7	2432_6740_18683_20633_20634_20635_20636	1	0
8	6740_18683_20633_20634_20635_20636_20637_20638	4	100
9	2432_4035_6740_18683_20633_20634_20635_20636_20638	1	0
9	4035_6740_18683_20633_20634_20635_20636_20637_20638	1	0
10	2432_4035_6740_18683_20633_20634_20635_20636_20637_20638	5	0

Come emerso precedentemente, la coppia sembra essere associata a basse frequenze di abbandono. Superati i 3 esami le alte percentuali di abbandono si associano a basse numerosità.

La scelta dei CdS

L'analisi di rete, introdotta nel successivo paragrafo, sarà condotta su un sottoinsieme limitato di CdS poiché, almeno in questo contesto, si predilige l'aspetto metodologico e l'operare su casi più semplici dal punto di vista dell'offerta formativa del primo anno.

I criteri nella scelta dei CdS sono finalizzati ad estrarre che permettono di creare un insieme sufficientemente omogeneo sia di studenti sia di insegnamenti da potere maggiormente attribuire l'evento abbandono al percorso dello studente durante il primo anno in termini quali e quanti esami ha superato.

In prima istanza si è deciso di selezionare delle lauree triennali che soddisfano i seguenti criteri:

- abbiano una numerosità di immatricolati sufficientemente alta da garantire l'applicabilità dei modelli statistici ed mostrare un numero accettabile di casi di abbandono. La numerosità minima è fissata in 150 matricole medie nel triennio
- la distribuzione per genere, tipo diploma e voto diploma degli immatricolati sia sufficientemente omogenea
- siano corsi triennali, dove il fenomeno dell'abbandono è più frequente
- abbiano un piano di studi al primo anno sufficientemente omogeneo tra loro in termini di SSD e CFU, possibilmente formato da soli esami obbligatori

Sono eleggibili 18 CdS:

Tab. 18 Eleggibilità dei CdS

CdS	Caratteristiche	Eleggibile
2120 - SCIENZE DELL'EDUCAZIONE - CLASSE L-19 (ACC.LIBERO)	4 curricula, ognuno con insegnamenti diversi per semestre	No
2099 - LETTERE - CLASSE L-10 (ACC.LIBERO)	2 curricula, ognuno con insegnamenti diversi	No
2121 - SCIENZE E TECNICHE PSICOLOGICHE - CLASSE L-24 (NUM.PROGR.)	1 curriculum, 3+5 insegnamenti, tutti obbligatori, no propedeuticità	Si
2113 - SCIENZE DELLA COMUNICAZIONE PER I MEDIA E LE ISTITUZIONI - CLASSE L-20 (ACC.LIBERO)	1 curriculum, 3+3 insegnamenti, tutti obbligatori, no propedeuticità però molti moduli, quindi più docenti, gestione dei codici insegnamento	Ni
2081 - ECONOMIA E FINANZA - CLASSE L-33 (ACC.LIBERO)	2 curricula ma uguali al primo anno, 4+4 tutti obbligatori, ci sono delle propedeuticità	Ni
2128 - SCIENZE POLITICHE E DELLE RELAZIONI INTERNAZIONALI - CLASSE L-36 (ACC.LIBERO)	2 curricula ognuno con insegnamenti diversi sia per cfu che per docenti titolari	No
2118 - SCIENZE DELLE ATTIVITÀ MOTORIE E SPORTIVE - CLASSE L-22 (NUM.PROGR.)	1 curriculum, 2+2(+1) insegnamenti (annuale), tutti obbligatori, ci sono propedeuticità, molti moduli, quindi più docenti gestione dei codici insegnamento	Ni
2079 - DISEGNO INDUSTRIALE - CLASSE L-4 (ACC.LIBERO)	Insegnamenti opzionali	No

CdS	Caratteristiche	Eleggibile
2172 - INFERMIERISTICA - CLASSE L/SNT1 (NUM.PROGR.)	1 curriculum 3+5 insegnamenti ma c'è il tirocinio, 2 moduli, propedeuticità	Ni
2198 - LINGUE E LETTERATURE - STUDI INTERCULTURALI - CLASSE L-12 (NUM.PROGR.)	Interclasse	No
2094 - INGEGNERIA GESTIONALE - CLASSE L-9 (NUM.PROGR.)	1 curriculum,4+3 insegnamenti, tutti obbligatori, no propedeuticità, no moduli	Si
2178 - INGEGNERIA INFORMATICA - CLASSE L-8 (NUM.PROGR.)	1 curriculum,2+2(+2) insegnamenti (annuali)tutti obbligatori, no propedeuticità 2 annuali, 2 moduli	Ni
2140 - INGEGNERIA BIOMEDICA - CLASSE L-9 (NUM.PROGR.)	2 sedi ma non è problema, 3 curriculum ma uguali al primo anno, no propedeuticità, 4+3(+1) insegnamenti obbligatori (annuale),1 modulo	Ni
2197 - LINGUE E LETTERATURE - STUDI INTERCULTURALI - CLASSE L-11 (NUM.PROGR.)	Interclasse	No
2112 - SCIENZE DEL TURISMO - CLASSE L-15 (ACC.LIBERO)	Insegnamenti opzionali	No
2194 - BENI CULTURALI: CONOSCENZA,GESTIONE,VALORIZZAZIONE - CLASSE L-1 (ACC.LIBERO)	2 curricula ognuno con insegnamenti diversi	No
2184 - STUDI FILOSOFICI E STORICI - CLASSE L-5 (ACC.LIBERO)	2 curricula ma uguali al primo anno, 4+4 insegnamenti tutti obbligatori, no propedeuticità	Si
2193 - INGEGNERIA DELL'INNOVAZIONE PER LE IMPRESE DIGITALI - CLASSE L-8 (ACC.LIBERO)	2 curricula ma uguali al primo anno, 4+4+1 insegnamenti (annuale) tutti obbligatori, un modulo no propedeuticità	Ni
SCIENZE BIOLOGICHE - L-13	È esposto al fenomeno degli abbandoni dovuti a successive immatricolazioni al CdLM in Medicina e Chirurgia	No

E' stata ulteriormente esclusa Scienze biologiche perché pur rientrando nei criteri è esposta al noto fenomeno degli abbandoni dovuti a successive immatricolazioni al CdLM in Medicina e Chirurgia, dopo il passaggio dei test di accesso.

Selezionati:

- ingegneria gestionale
- scienze e tecniche psicologiche
- studi filosofici e storici

Ulteriormente selezionabili (in parentesi perché non scelti in prima istanza)

- economia e finanza (per propedeuticità)

- infermieristica (per propedeuticità, moduli e tirocinio)
- ingegneria biomedica (per modulo)
- ingegneria dell'innovazione per le imprese digitali (per moduli)
- ingegneria informatica (per annualità e moduli)
- scienze della comunicazione (per moduli)
- scienze motorie (per moduli)

Tab. 19 distribuzioni di alcune caratteristiche dei CdS selezionati

CdS	N° immatricolati nel triennio	% femmine	% Tipo diploma	Voto medio diploma
2094 - Ingegneria gestionale	761	31.9	Liceo scientifico: 71.6 Liceo Classico: 12.1	85.4
2121- Scienze e tecniche psicologiche	1499	77.6	Liceo Classico: 26.4 Liceo scientifico: 22.4 Istituto Magistrale: 21.2	79.9
2184 - Studi filosofici e storici	681	48.3	Liceo Classico: 29.1 Liceo scientifico: 23.2 Istituto Magistrale: 14.4	79.7

Nonostante il superamento dei suddetti criteri, restano alcuni ulteriori problemi di difficile gestione:

- il possibile cambiamento di insegnamenti di anno in anno, il cambiamento di ordinamento o il semestre di erogazione insegnamenti pongono un limite alla longitudinalità delle osservazioni
- alcuni CdS utilizzano le cosiddette materie opzionali a gruppi, gruppi di insegnamenti tra cui ogni studente può scegliere un numero predeterminato di CFU che rendono disomogenei i piani di studio individuali e generano variabilità tra frequenze di superamento dell'esame. Tipicamente questi insegnamenti hanno frequenze di superamento minori poiché non sono obbligatori.

ANALISI DELLE RETI

In questo capitolo si introducono i concetti basilari dell'analisi di rete adeguati al contesto in cui ci muoviamo, ovvero la relazione tra studente e gli esami superati al primo anno e l'associazione tra questa relazione e l'abbandono o la prosecuzione al secondo.

La rete bipartita

La rete più adatta a modellare il percorso del primo anno è la **rete bipartita studente/esame**, in cui una forma di nodo rappresenta lo studente e un'altra forma di nodo rappresenta l'esame e il link tra un nodo studente e un nodo esame rappresenta l'aver sostenuto quell'esame e acquisito i relativi CFU. Coerentemente, la rete bipartita ammette solo link tra le due forme di nodi e non tra nodi della stessa forma. La caratteristica del proseguimento/abbandono è evidenziata dal colore del nodo studente (colore verde=prosecuzione rosso=abbandono)

Attributi della rete:

- bipartita, indiretta;
- non pesata, ma può essere pesata se il peso dei link è dato dai CFU dell'esame;
- per dare evidenza all'aspetto dinamico dell'ottenimento dei CFU, la rete può essere formata da 2 a 4 layer: primo e secondo semestre, o primo semestre, sessione intra secondo semestre, secondo semestre, settembre. In prima istanza si è scelto di operare con 2 layer.

Attributi del nodo studente:

- colore: prosecuzione (verde), abbandono (rosso);
- forma: identifica il nodo studente (cerchio);
- dimensione: proporzionale al n° esami sostenuti dallo studente (o al n° cfu acquisiti);
- label: identificativo anonimo dello studente.

Attributi del nodo insegnamento:

- colore: semestre di erogazione dell'insegnamento;
- forma: identifica il nodo insegnamento (quadrato);
- dimensione: proporzionale ai CFU;
- label: identificativo dell'insegnamento.

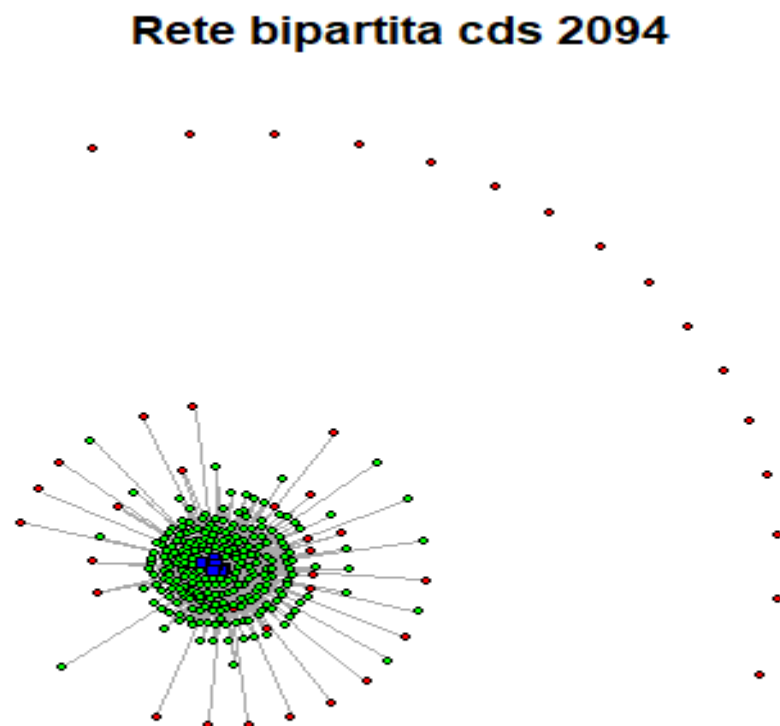
Attributi del link:

- tipo: indiretto.

Caratteristiche della rete:

- grado del nodo: quanti esami ha sostenuto (o quanti cfu ha acquisito);
- il layout maggiormente utilizzato è di fruchterman e reingold si interpreta come segue:
 - i nodi quadrati rappresentano gli insegnamenti, i nodi circolari gli studenti;
 - i link associano gli studenti agli esami superati;
 - al centro ci sono i noi di grado maggiore, ovvero gli insegnamenti di colore blu. Il grado di ogni insegnamento è il numero di studenti che ne hanno superato l'esame (nell'ordine delle decine, mentre i nodi studente hanno grado nell'ordine delle unità);
 - dall'interno verso l'esterno, i nodi studente sono disposti in anelli concentrici rispetto ai nodi esame, si posizionano circolarmente i nodi di pari grado ovvero gli studenti che hanno superato tutti gli esami, con distanza crescente al decrescere del grado (studenti con meno esami sono disposti verso l'esterno fino a coloro che non hanno superato alcun esame, cioè i nodi isolati);
 - il colore dei nodi studenti identifica l'evento finale, in verde le prosecuzioni, in rosso gli abbandoni.

Fig. 1 Rete bipartita³



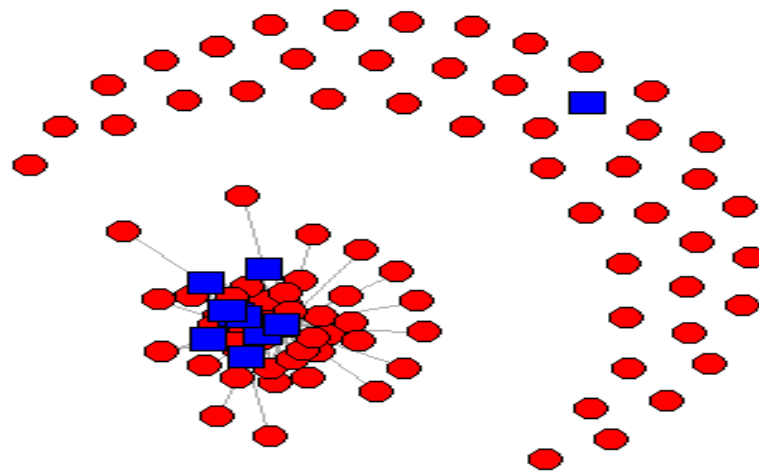
³ Tutte le reti sono state costruite e rappresentate con il pacchetto igraph⁽⁴⁾, con l'ausilio del framework tidyverse⁽¹¹⁾, all'interno della suite R⁽⁹⁾

Si legge chiaramente come tutti gli studenti che non hanno superato alcun esame, i nodi isolati, abbiano abbandonato, nodi rossi, sebbene siano in numero limitato. Al contrario coloro che hanno superato molto o anche tutti gli esami previsti, i nodi studente più vicini ai nodi esame, abbiano poi proseguito, i nodi verdi, e sono in gran numero.

Una prima operazione applicabile alla rete bipartita è l'estrazione di una **sottorete bipartita indotta**, un sottoinsieme di nodi e di link che li uniscono, di particolare interesse. Si è scelto di estrarre la sottorete degli abbandoni. Si noti l'ulteriore informazione ricavata non derivabile dalla rete bipartita: un esame non è stato superato da nessuno studente che abbia poi abbandonato.

Fig. 2 Sottorete bipartita degli abbandoni

Sottorete bipartita degli abbandoni, cds: 2121

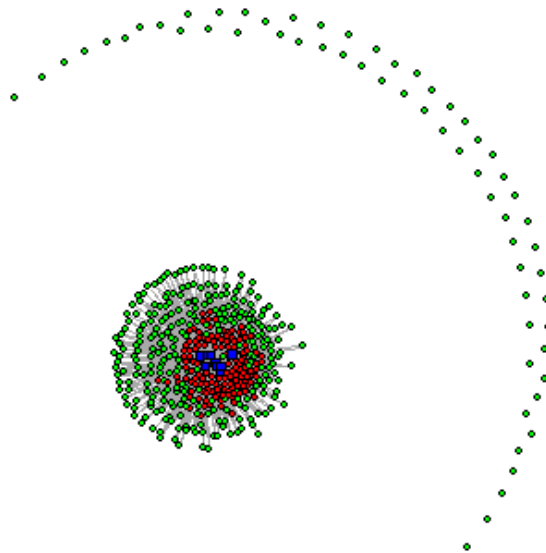


Una ulteriore operazione è la costruzione della **rete bipartita complementare** ovvero la rete che ha link tra nodi dove la rete originaria non li ha e viceversa.

Si noti come emerge ancora più chiaramente l'associazione tra abbandono ed esami non superati: i nodi rossi stanno tutti al centro della rete.

Fig. 3 Rete bipartita complementare

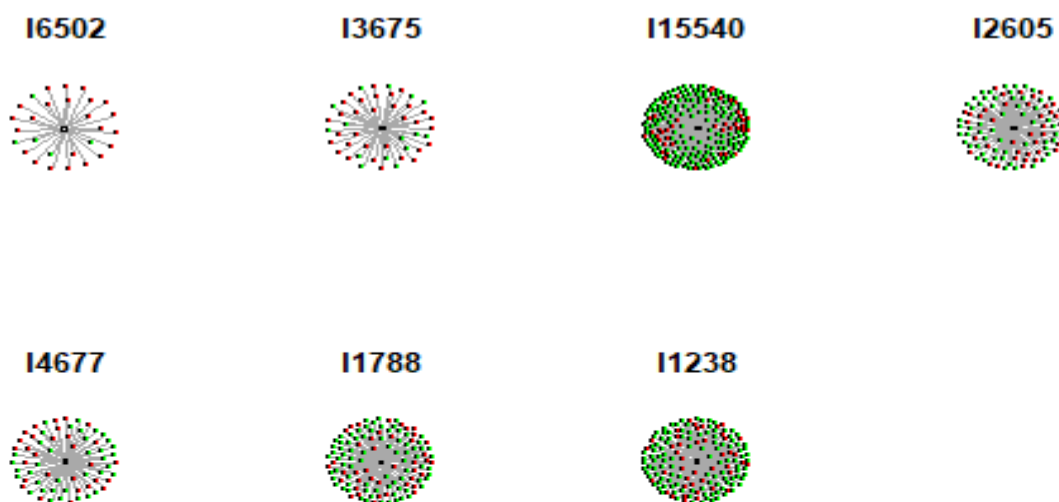
rete complementare cds: 2184



Altra operazione sulla rete bipartita è l'estrazione delle "reti ego" ossia di sottoreti bipartite formate da un esame e da tutti gli studenti che lo hanno superato o, come nella prossima figura, che non hanno superato quell'esame, eventualmente colorati in funzione dell'abbandono.

Si noti come la prima rete ego in alto a destra mostri pochi link ma quasi verso nodi rossi, quindi pochi studenti non hanno superato l'esame ma proprio quelli che hanno poi abbandonato.

Fig. 4 Ego Rete bipartita complementare



Reti Ego indotti dalla rete complementare, cds 2094

Infine, una operazione molto importante sulla rete bipartita è l'estrazione delle due proiezioni: la **proiezione sugli studenti** e la **proiezione sugli esami**. La prima è formata da tutti i nodi studente e dai link tra essi dove un link tra due nodi è stabilito se i due studenti condividono almeno un insegnamento. E' una rete pesata, il peso è dato dal numero di esami superati da entrambi gli studenti. La seconda è formata da tutti i nodi esame e dai link tra essi dove un link tra due nodi è stabilito se almeno uno studente ha superato entrambi gli esami. E' una rete pesata, il peso è dato dal numero di studenti che ha superato una coppia di esami.

La proiezione sugli studenti

E' una rete indiretta pesata in cui ogni nodo rappresenta uno studente e il link è definito tra studenti che hanno sostenuto almeno uno stesso esame. Il peso del link è definito dal numero di esami condivisi dei due studenti.

Attributi del nodo:

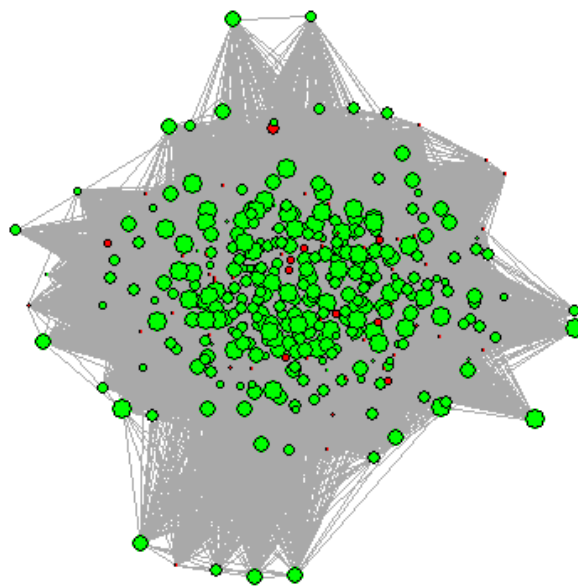
- colore: prosecuzione/abbandono;
- dimensione: proporzionale al n° esami;

Attributi del link:

- dimensione: proporzionale al n° esami acquisiti assieme

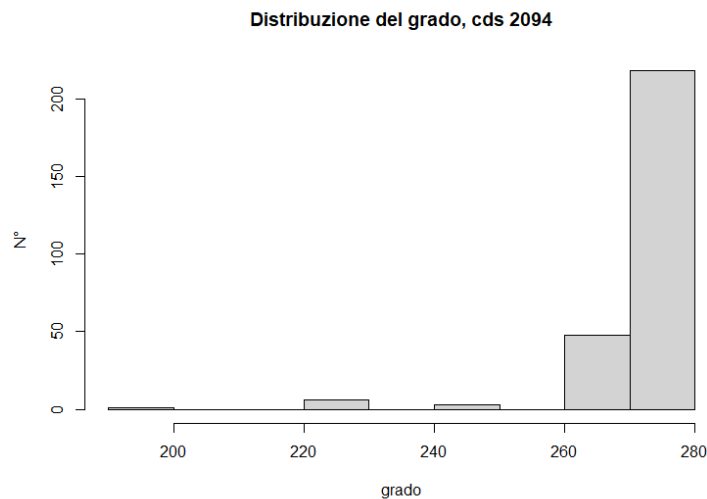
E' una rete molto densa. Ad esempio la proiezione sugli studenti del CdS 2094 ha 276 nodi (al netto degli studenti a 0 esami) e 37479 link, pari a un densità di 0.9875. La rete bipartita originaria ha una densità di 0.0166.

Fig. 5 Rete proiezione sugli studenti della rete bipartita

proiezione studenti cds: 2094

Di seguito l'istogramma del grado dei nodi, da cui si evince come la quasi totalità dei nodi abbia più di 200 link.

Fig. 6 Distribuzione del grado dei nodi della proiezione studenti



Verificare la necessità di link filtering.

Ego network:

- di uno studente, tutti gli studenti che condividono almeno uno dei suoi esami, quindi sia studenti con meno esami che con più esami di quelli da lui sostenuti
- di un esame, tutti gli esami che l'hanno superato, di più soprattutto l'insieme degli studenti che li hanno superati congiuntamente

La rilevazione delle comunità sulla proiezione degli studenti

All'interno dell'analisi di rete una tecnica molto diffusa è la ricerca delle cosiddette comunità, ossia di insiemi di nodi che condividono caratteristiche o, ancora meglio, relazioni. Informalmente una comunità è un insieme di nodi in cui il numero di link fra i nodi appartenenti alla stessa comunità è maggiore del numero di link tra nodi esterni alla comunità, una proprietà formalizzata nel concetto di modularità (2) (3) (4).

Esistono diversi algoritmi che dato in ingresso una rete restituiscono le comunità e l'appartenenza di ogni nodo ad una comunità. Si è scelto il `cluster_fast_greedy`, basato appunto sulla modularità e successivamente si è calcolato il numero di prosecuzioni, di abbandoni e la percentuale di abbandono per ogni comunità rilevata sui 3 CdS.

Tab. 20 Associazione tra comunità rilevate e abbandono

CdS 2094				
Id Comunità	N° prosezioni	N° abbandoni	% Abbandono	N° medio di esami
0 (nessun esame)	0	16	100,0	0
1	12	0	0,0	4.7
2	137	0	0,0	6.2
3	100	27	21,3	3.3
CdS 2121				
Id Comunità	N° prosezioni	N° abbandoni	% Abbandono	N° medio di esami
0 (nessun esame)	12	44	78,6	0
1	155	3	1,9	7.0
2	161	5	3,0	6.4
3	153	37	19,5	4.0
CdS 2184				
Id Comunità	N° prosezioni	N° abbandoni	% Abbandono	N° medio di esami
0 (nessun esame)	11	93	89,4	0
1	45	45	50,0	2.8
2	110	8	6,8	5.6
3	164	18	9,9	5.6

La procedura ha restituito 3 comunità per ciascuno dei CdS. La capacità predittiva è scontata per la comunità 0, quella formata dei nodi isolati (per cui si è creato un esame fittizio a cui collegarli preventivamente), e appare sufficientemente discriminate per le altre, si possono infatti identificare comunità a basso tasso di abbandono (comunità di successo) e ad alto tasso di abbandono (comunità di insuccesso).

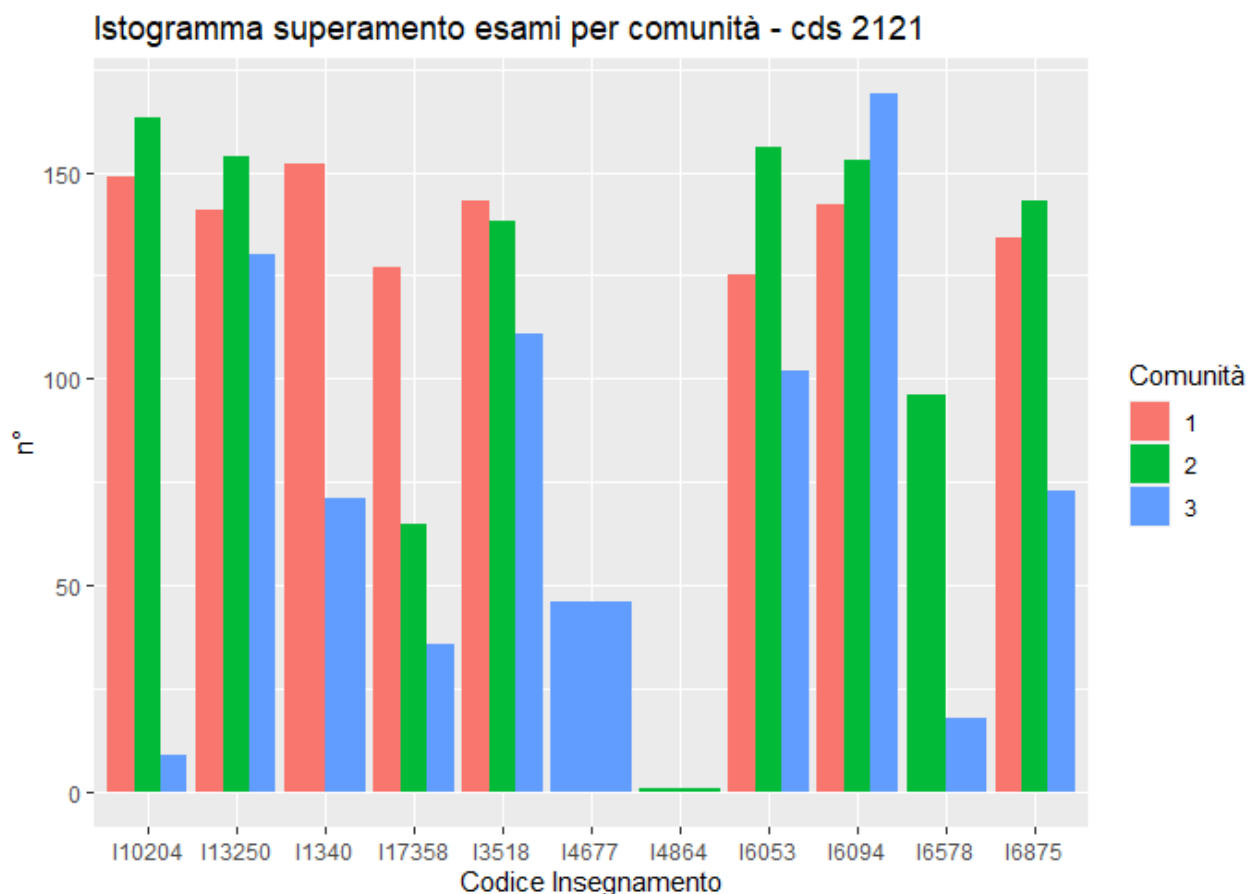
Il CdS 2904 mostra una piccola comunità, la 1 con 12 studenti, con nessun abbandono, due comunità più grandi, oltre 100 studenti, con tassi di abbandono molto diversi (0% vs 21.3%), caratterizzate da un n° medio di esami superati altrettanto diverso (6.2 vs 3.3)

I CdS 2121 mostra 3 comunità egualmente numerose in cui le prime 2 comunità con bassi tassi di abbandono e alto n° medio di esami e la terza con un alto di abbandono e un basso n° medio di esami.

Anche il CdS 2184 mostra una comunità ad alto tasso di abbandono, la 1, caratterizzata da un numero medio di esami basso rispetto alle altre due.

E' confermata dunque la soglia dei 4 esami, rilevata nell'analisi descrittiva, al di sotto della quale le frequenze di abbandono all'interno delle comunità sono più alte.

Fig. 7 Frequenze assolute superamento degli esami condizionato distinte per comunità



Le 3 comunità sono di ampiezza simile. Dalla comunità 3, quella a maggior tasso di abbandono, si possono estrarre insegnamenti sovraespressi, ovvero a maggiore frequenza, ad esempio I4677 o I6094, e sottoespressi, I10204, I17358 e I4864. Si noti come il CdS 2121 abbia un numero medio di esami più alto degli altri CdS, a prescindere dalla divisione in comunità, per cui in questo caso si potrebbe valutare l'innalzamento della soglia di sicurezza degli esami superati da 4 a 5. Inoltre, alla luce di un alto tasso di abbandono all'interno di una comunità con un così alto numero di esami superati, andrebbe anche valutato se almeno parte degli abbandoni non siano in realtà dei passaggi di CdS.

La comunità della proiezione degli studenti

Una clique è definita come una sottorete completa, ovvero dove ogni nodo è connesso con tutti gli altri nodi della clique.

Dalla proiezione degli studenti si possono estrarre diverse cliques di interesse. Ad esempio, l'insieme degli studenti che superano esattamente gli stessi N esami formano una clique.

Da questa clique si diramano altri link esterni che connettono ogni studente della clique con ognuno degli altri studenti con cui condividono almeno un esame insieme, quindi in definitiva con altre cliques. Dato che un sottoinsieme di una clique è una clique a sua volta.

Dunque, fissata una clique Le più importanti sono le cliques di N-1 esami sottoinsieme proprio della clique data,

Ad esempio la clique degli studenti che hanno superato Matematica 1-Statistica 1- informatica-calcolo delle probabilità. Le tre 3 cliques sono gli insiemi degli studenti che anno superato (Matematica 1-Statistica 1- informatica) o (Matematica 1-Statistica 1- calcolo delle probabilità) o (Matematica 1-informatica-calcolo delle probabilità) o (Statistica 1- informatica-calcolo delle probabilità) . Ognuno degli studenti della clique è connesso a tutti gli studenti di una dell

Attributi del nodo:

- colore: verde = prosecuzione, rosso = abbandono
- dimensione: proporzionale al numero di esami superati
- label: identificativo dello studente

Attributi del link:

- dimensione: proporzionale al n° di coppie di studenti che lo hanno sostenuto
- tipo: non diretto

cliques: si estraggono col peso

La proiezione sugli insegnamenti

E' una rete indiretta pesata in cui ogni nodo rappresenta un insegnamento e il link è definito tra insegnamenti sostenuti almeno uno stesso studente. Il peso è rappresentato dal numero di studenti che hanno condiviso i due esami.

Attributi del nodo:

- colore: l'ambito dell'insegnamento: A = Attività di Base, B = Caratterizzante, C = Affine Integrative, D = Scelta Studente, E = Altre attività
- dimensione: proporzionale al grado
- label: identificativo dell'insegnamento

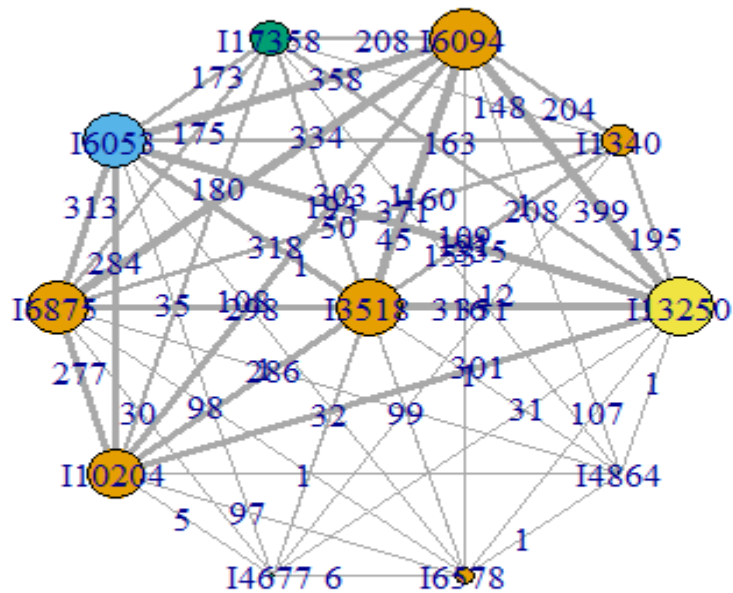
Attributi del link:

- tipo: non diretto
- dimensione: proporzionale al n° di coppie di studenti che lo hanno sostenuto

di seguito la proiezione degli insegnamenti del cds 2121.

Fig. 8 Rete proiezione sugli insegnamenti della rete bipartita

Proiezione insegnamenti, cds: 2121

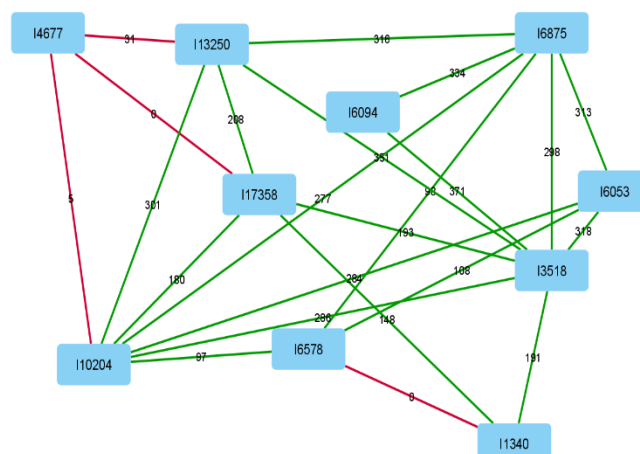


La rete è tipicamente completamente connessa. Si perde il riferimento all'abbandono.

Al fine di testare se le frequenze congiunte empiriche sono compatibili con un modello casuale, oppure indicano associazioni significative tra gli esami sostenuti, si è applicata la tecnica delle reti statisticamente validate ⁽¹⁰⁾. Il modello di urna è quello ipergeometrico, a cui è stata applicata una correzione di Bonferroni al 5% per test multipli. I risultati, presentati nella seguente tabella, indicano le correlazioni statisticamente significative con i relativi p-value, dunque definiscono le reti di Bonferroni. I link associati a correlazioni negative tendono a separare i nodi ad essi adiacenti, come mostra la successiva figura.

Fig. 9 Rete proiezione sugli insegnamenti statisticamente validata

rete proiezione sugli insegnamenti statisticamente validata, cds 2121



Tab. 21 Link statisticamente validati della rete sulla proiezione sugli insegnamenti

Ins_1	Ins_2	N° stud Ins1	N° stud Ins2	N° stud Ins1 e Ins2	N° Stud tot	Corr	p-value right tail	p-value left tail	Dummy link positivo-negativo
I10204	I13250	320	424	301	490	0,30	0,0000	1,0000	1
I10204	I17358	320	226	180	490	0,28	0,0000	1,0000	1
I10204	I3518	320	388	286	490	0,34	0,0000	1,0000	1
I10204	I4677	320	45	5	490	-0,36	1,0000	0,0000	-1
I10204	I6053	320	381	284	490	0,36	0,0000	1,0000	1
I10204	I6578	320	114	97	490	0,23	0,0000	1,0000	1
I10204	I6875	320	350	277	490	0,46	0,0000	1,0000	1
I13250	I17358	424	226	208	490	0,15	0,0006	0,9998	1
I13250	I3518	424	388	351	490	0,22	0,0000	1,0000	1
I13250	I4677	424	45	31	490	-0,16	0,9997	0,0010	-1
I13250	I6875	424	350	316	490	0,17	0,0002	0,9999	1
I1340	I17358	221	226	148	490	0,38	0,0000	1,0000	1
I1340	I3518	221	388	191	490	0,16	0,0002	0,9999	1
I1340	I6578	221	114	0	490	-0,50	1,0000	0,0000	-1
I17358	I3518	226	388	193	490	0,14	0,0011	0,9995	1
I17358	I4677	226	45	0	490	-0,29	1,0000	0,0000	-1
I3518	I6053	388	381	318	490	0,20	0,0000	1,0000	1
I3518	I6094	388	453	371	490	0,23	0,0000	1,0000	1
I3518	I6875	388	350	298	490	0,23	0,0000	1,0000	1
I6053	I6578	381	114	108	490	0,22	0,0000	1,0000	1
I6053	I6875	381	350	313	490	0,44	0,0000	1,0000	1
I6094	I6875	453	350	334	490	0,18	0,0002	1,0000	1
I6578	I6875	114	350	98	490	0,18	0,0000	1,0000	1

Elementi di machine learning

Le reti fin qui definite, bipartita, complementare, ego, proiezioni e quelle relative all'analisi dei flussi, motifs compresi, sono dinamiche.

Le reti si aggiornano automaticamente con periodicità fissata:

- le coorti degli immatricolati sono aggiornate annualmente;
- l'offerta formativa del primo anno è aggiornata semestralmente;
- il superamento degli esami è aggiornato di sessione in sessione;
- l'evoluzione dei motifs è aggiornato di sessione in sessione;
- le probabilità di abbandono sono aggiornate annualmente.

ANALISI DEI PERCORSI

L'analisi di rete prevede **lo studio dei percorsi** e degli indici associati (shortest path, closeness, etc). Nella carriera studentesca del primo anno, il percorso è formato dall'insieme degli esami superati nelle successive sessioni di esami fino all'esito finale, prosecuzione o abbandono.

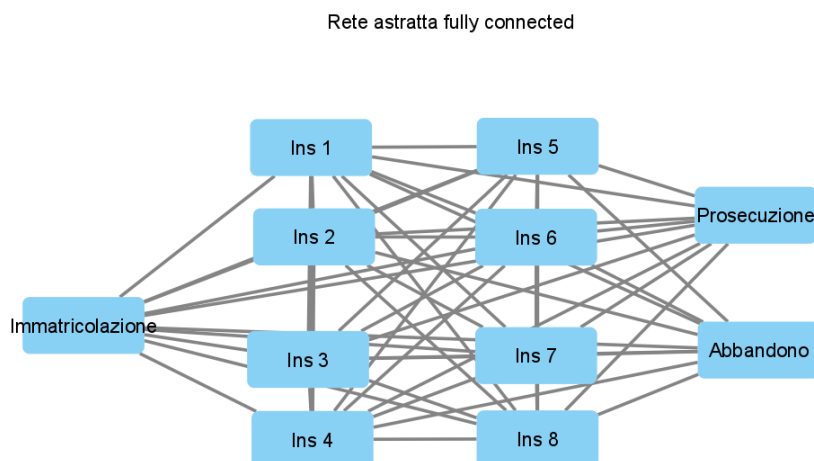
Costruiamo un esempio con 4 insegnamenti a semestre, in assenza di propedeuticità.

- **layer 0, immatricolazione**, nessun esame studente appena immatricolato; i link rappresentano tutti i possibili “primo esame” o direttamente “prosecuzione/abbandono” se lo studente non supera alcun esame alla fine del primo anno;
- **layer 1, primo semestre**, 4 nodi in colonna, sono i 4 insegnamenti erogati nel primo semestre; i link rappresentano tutti i possibili percorsi dello studente in qualsiasi ordine; in qualsiasi ordine significa che ogni esame può essere stato superato o al primo semestre, come primo esame o successivo, o al secondo semestre anche dopo aver superato un esame erogato nel secondo semestre, dunque “tornando indietro” da un nodo del layer 2; se l'ultimo esame superato appartiene a questo semestre il nodo è direttamente connesso allo status finale;
- **layer 2**: sono i 4 insegnamenti erogati nel **secondo semestre**. i link rappresentano tutti i possibili percorsi dello studente in qualsiasi ordine; i link si interpretano come appena descritto;
- **layer 3**: due nodi di altra natura. Nodo prosecuzione (P) e nodo abbandono (A). Ovvero **lo status** dell'immatricolato **a fine anno**.

E' una rete **completamente connessa**⁴: dalla condizione iniziale di immatricolato si può andare direttamente a quella finale di abbandono, o prosecuzione, oppure verso qualsiasi insegnamento erogato del primo o del secondo; da ognuno dei nodi del primo semestre può essere direttamente connesso a ognuno degli altri insegnamenti del primo semestre o a ognuno degli insegnamenti del secondo semestre, o ai nodi finali P e A. Ed ognuno dei nodi del secondo semestre può essere direttamente connesso a ognuno degli altri insegnamenti del primo del secondo semestre, o ai nodi finali P e A. Il nodo iniziale può connettersi a qualunque nodi del primo o del secondo o anche a I nodi finali P e A non possono essere connessi tra loro.

⁴ Nel nostro caso la rete sarà non orientata. Se si volesse tenere conto dell'ordine temporale del superamento degli esami, la rete sarebbe orientata. Il nodo immatricolazione può solo emettere link (Imm->); i nodi finali possono solo ricevere (->Abb/Pros); mentre gli altri nodi possono sia emettere che ricevere (->InsX->). La condizione di completezza deriva dalla mancanza di propedeuticità.

Fig. 10 Rete astratta completamente connessa



La rete può descrivere **due distinte tipologie di percorsi**:

- il percorso **individuale**, l'insieme di esami superati da un singolo studente;
- il percorso **aggregato** che sintetizza tutti gli studenti con lo stesso percorso, tipicamente attraverso l'attributo size dei link.

Dal punto di vista delle **tipologie di analisi** possiamo distinguere:

- **descrittiva**: descrizione dei percorsi individuali o aggregati tra l'immatricolazione e la prosecuzione o l'abbandono;
- **previsiva**: si valuta la probabilità di abbandono condizionatamente al percorso osservato basandosi sui percorsi conclusi negli anni precedenti in modo da potere intervenire in itinere. Più chiaramente: si fissa il percorso di uno studente ad esempio al primo semestre, si estraggono tutti i percorsi degli studenti degli anni passati con il medesimo percorso al primo semestre e si calcola la frequenza media di abbandono;
- **di intervento**: alla luce della precedente analisi, si può suggerire il percorso del secondo semestre orientato alla minimizzazione della probabilità di abbandono. Inoltre si possono valutare le probabilità di abbandono togliendo esami ad un determinato percorso osservato al fine di valutare l'apporto di ogni singolo esame alla probabilità di abbandono;
- **retrospettiva**: a fine anno e a condizione finale nota, si può valutare cosa sarebbe potuto accadere se uno studente avesse intrapreso un percorso in parte diverso da quello seguito (ad

esempio cambiato un esame con un altro), ovvero valutare come passare da un percorso ad alto rischio abbandono in un percorso a basso rischio (e viceversa).

La prima tipologia è statica, si guarda a ciò che è stato fino al momento dell'osservazione. In parte anche la seconda: il primo semestre è osservato mentre il secondo semestre è stimato dagli anni precedenti. La terza tipologia è dinamica, si devono fare ipotesi sul percorso futuro. La quarta inizia dallo stato finale dello studente, torna indietro di uno o più esami, li sostituisce con altri e ritorna allo stato finale opposto a quello di partenza.

Infine, dal punto di vista dei **portatori di interesse** possiamo distinguere:

- **Lo studente** può essere interessato al proprio e **percorso individuale**, a quello **che minimizzi la probabilità di abbandono**, condizionatamente al tragitto già attraversato e alla luce dei percorsi aggregati dei suoi colleghi di coorte (allo scopo di posizionarsi) e delle coorti precedenti (allo scopo di valutare le probabilità di prosecuzione o abbandono).
- **Il docente** del corso può essere interessato alla probabilità di abbandono associata al suo insegnamento o al **posizionamento dell'insegnamento** nei percorsi degli studenti (quali sono gli altri insegnamenti a cui più si associa il proprio in termini di).
- **Il coordinatore** può essere interessato a percorsi aggregati, a valutare quali **insegnamenti siano maggiormente associati a maggiore rischio di abbandono**, in modo da potere intervenire, se necessario, sugli studenti, con delle azioni di orientamento, sui docenti, per valutare la persistenza di **“materie scoglio”** o infine alla revisione dell'offerta formativa del primo anno.

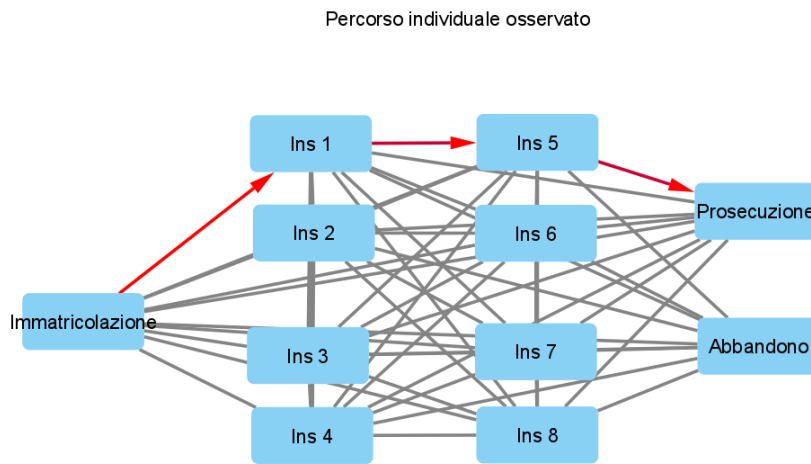
L'analisi osservazionale: i percorsi del primo anno

E' la rete descrittiva dei percorsi osservati, individuali o aggregati.

La rete individuale

E' il percorso osservato durante l'anno corrente e rappresenta l'insieme ordinato temporalmente degli esami sostenuti dallo studente. Può essere costruito sia su **percorsi dell'anno corrente** sia su **percorsi dei precedenti anni**.

Fig. 11 Esempio di percorso individuale osservato



Si noti come il percorso osservato si traduce in una rete ordinata, immatricolazione → Ins 1 → Ins 5 → Prosecuzione, ma non necessariamente uguale a quello che appare nella rete, cioè Ins 1 è stato superato prima di Ins 2, è compatibile con 2 scenari diversi, Ins 1 superato al primo semestre e Ins 2 al secondo; o entrambi superati al secondo semestre ma Ins 1 in data anteriore a Ins 2. E' incompatibile con l'aver superato Ins 5 al primo semestre poiché anteriore al semestre di erogazione. Inoltre lo stesso percorso avrebbe potuto avere un ordine diverso immatricolazione → Ins 5 → Ins 1 → Prosecuzione, ovvero un percorso all'indietro dal secondo semestre (di erogazione) al primo. Questo è possibile solo se entrambi gli esami sono sostenuti a secondo semestre.

Il percorso osservato dello studente è rappresentato sulla rete colorando e orientando il suo tragitto. Se si vuole conservare l'aspetto temporale si usano le frecce. Se non si vuole conservare l'aspetto temporale si usano gli archi: Il caso più semplice è l'abbandono, o la prosecuzione, senza esami, ossia una freccia dall'immatricolazione a A o P. Segue il caso con un solo esame, o al primo o al secondo semestre. Poi il caso con due esami, entrambi al primo semestre o entrambi al secondo oppure il primo al primo semestre e il secondo al secondo o viceversa. Chiaramente il numero di combinazioni esplode all'aumentare degli esami.

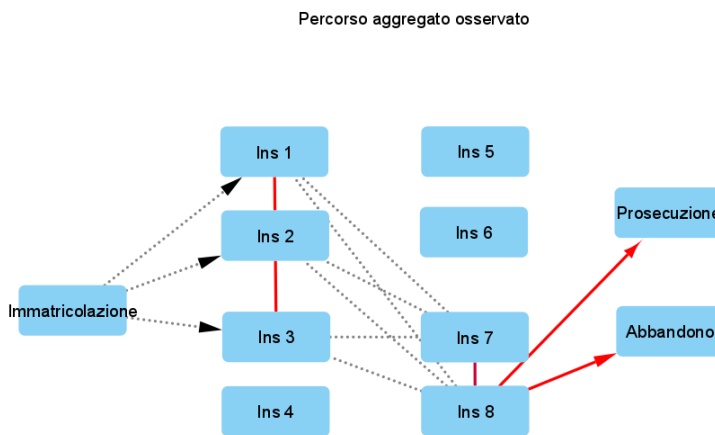
La rete individuale aggregata

E' uguale a alla rete individuale con due importanti **eccezioni**:

- mentre un percorso individuale termina con uno solo dei due output possibili, prosecuzione o abbandono, la **rete aggregata contempla entrambi gli output**, poiché a parità di percorso l'esito finale può cambiare da studente a studente
- È una rete pesata in cui i **pesi dei percorsi** sono proporzionali alle frequenze degli studenti che li percorrono

- Pur non essendo obbligati, si è scelto di configurare la rete come “non diretta” ovvero di **non tenere conto ordine temporale del superamento degli esami**, primariamente perché il calendario degli esami cambia di anno in anno e dunque l’ordine temporale del superamento degli esami può non dipende dallo studente, e in secondo luogo perché tenere conto del tempo aumenta il numero dei percorsi (teorici e osservati) e ne diminuisce la frequenza osservata rispetto al percorso aggregato (banalmente 10 studenti hanno superato gli esami Ins1 e Ins2, un percorso, ma 7 in quest’ordine e 3 nell’opposto, due percorsi).

Fig. 12 Esempio di percorso aggregato osservato



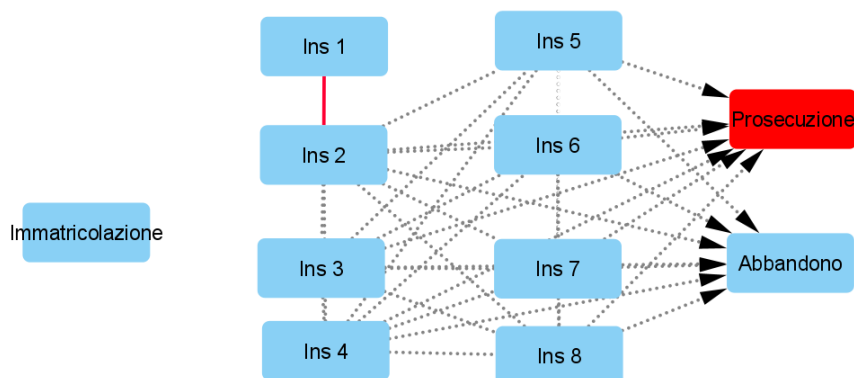
La rete aggregata si interpreta come segue: le frecce nere discontinue rappresentano i possibili ordinamenti temporali del superamento degli esami di cui si è deciso di non tenere conto. Restano i percorsi rossi non orientati che fissano gli esami superati (indipendentemente dall’ordine temporale all’interno del semestre) ed entrambi gli esiti finali possibili, in proporzione variabile in funzione quali-quantitativa degli esami precedentemente superati.

Oltre ad potere essere analizzata per se, è uno strumento indispensabile per le **analisi previsionali**, di intervento e retrospettive poiché offre degli scenari di **percorsi aggregati osservati di cui si conosce l’esito finale**.

Fig. 13 Esempio di percorso previsionale aggregato

Rete previsionale - probabilità di di abbandono o prosecuzione dati 2 esami al primo semestre

E' evidente la marginalizzazione rispetto a tutti i possibili percorsi osservati



In questo esempio lo studente ha superato due esami al primo semestre e si vuole calcolare la **probabilità previsiva dell'abbandono** $P(\text{abbandono} \mid \text{percorso al } 1^{\circ} \text{ semestre})$, a partire dai percorsi aggregati delle coorti precedenti che condividono il suo stesso percorso al primo semestre. E' evidente la marginalizzazione rispetto sia a tutti i possibili percorsi del secondo semestre, sia al sottoinsieme di quelli eventualmente osservati negli anni precedenti, di cui è noto l'esito finale.

Ma proprio perché è noto l'esito finale è possibile calcolare la **frequenza diagnostica** di avere o non avere sostenuto un determinato esame $P(\text{avere/non avere sostenuto l'esame } X \mid \text{abbandono Ins1,Ins2})$ o generalizzando la frequenza di un percorso al secondo semestre $F(\text{percorso al } 2^{\circ} \text{ semestre} \mid \text{abbandono, Ins1,Ins2})$.

Intervento: uno strumento del COT per l'orientamento in itinere

L'analisi descrittiva ha mostrato come si possa fissare una soglia di 4 esami raggiunti i quali la frequenza di abbandono si attesta a livelli accettabili (7%) e dunque l'analisi della variabilità introdotta dal considerare quali esami, oltre a quanti, si associano alle frequenze di abbandono ha poco valore operativo. La stessa analisi ha mostrato come l'inattività didattica al secondo semestre sia più rischiosa di quella al primo semestre. Si riporta di seguito parte della tabella Tab. 10

Tab. 22 Distribuzioni doppie assolute e percentuali del numeri studenti condizionato al superamento di 4 esami al variare del numero al primo e al secondo semestre e dell'abbandono condizionato al variare numero di esami superati al primo e al secondo semestre

N° studenti	%° studenti	N° esami I semestre	N° esami II semestre	N° abbandoni	% abbandoni
221	0,8	0	4+	8	3,6
877	3,0	1	3	34	3,9
1713	5,9	2	2	101	5,9
721	2,5	3	1	79	11
159	0,5	4+	0	47	29,6

E' indubbio che lo **sforzo** deve essere diretto nel **monitoraggio a fine primo semestre e nell'intervento durante il secondo** allo scopo di minimizzare la probabilità di abbandono.

La struttura dell'Ateneo preposta alle attività di tutoraggio in itinere è il Centro di Orientamento e Tutorato (COT). Si potrebbe rafforzare l'azione del COT con un applicativo che indirizzi il personale nei suggerimenti agli studenti. Lo scenario è il seguente: lo studente si presenta allo sportello, spontaneamente o perché sollecitato dall'Ateneo stesso, e la conversazione focalizza su quali siano i prossimi insegnamenti da seguire o esami da superare, se ad esempio sia più utile recuperare un esame del primo semestre o sostenere qualcuno del secondo, o consigliare decisamente tutto il percorso del secondo semestre.

A tal fine bisogna estrarre i percorsi aggregati del primo semestre uguali a quello dello studente che si sono conclusi con una frequenza la prosecuzione accettabile (approccio empirico), ove ve ne siano e propoli allo studente.

Nel linguaggio delle reti, i **motifs** sono definiti come piccoli **sottografi ricorrenti in reti più ampie** ⁽⁸⁾. Nel nostro caso un motif è l'insieme degli esami superati al primo semestre che ricorre nell'insieme più ampio degli esami a fine anno.

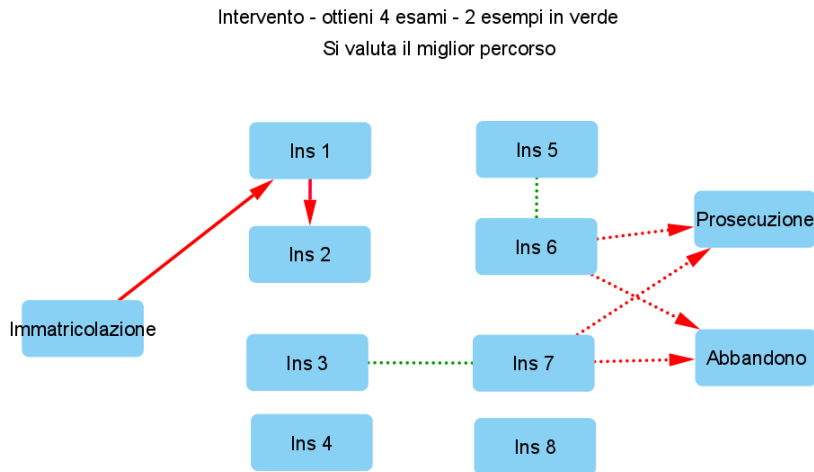
Nessun esame al primo semestre. Suggestire un percorso di 4 esami dopo un semestre bianco potrebbe rivelarsi poco credibile. In questo casi una soluzione naif potrebbe essere quella di suggestire l'esame o la coppia di esami frequenti, senza legarli a probabilità di abbandono.

Un esempio empirico tratto dal CdS 2184: esami (marginalmente) più frequenti: I4062, 299 esami, I340, 261 esami.

Coppie di esami (marginalmente) più frequenti (dalla proiezione sugli insegnamenti): 1 (I4062-I19192), 236 coppie di studenti, (I4062-I19234) 230 coppie di studenti, (I4062-I16585) 228 coppie di studenti.

Se lo studente ha superato almeno 2 esami è ipotizzabile suggestire ulteriori esami legandoli alle frequenze di abbandono.

Fig. 14 Esempio di intervento finalizzato al percorso del secondo semestre



Un esempio empirico tratto dal CdS 2094:

Motif iniziale: "3675_6502"

N° studenti della coorte con quel motif: 6

Tab. 23 Motifs allargato a 3: suggerire ulteriore esame al 2° semestre

Ulteriore esame	N° studenti coorti precedenti	N° abbandoni
15540	1	0
1788	2	0
2605	4	1

Il percorso consigliato di un ulteriore esame è il "1788".

Tab. 24 Motifs allargato a 4: minimizzare la probabilità di abbandono

Ulteriori esami	N° studenti	N° abbandoni
1788_1238	1	1
2605_1788	7	0
4677_2605	1	2

Dato che nessuno studente ha abbandonato, si può consigliare il percorso più frequente: "2605_1788"

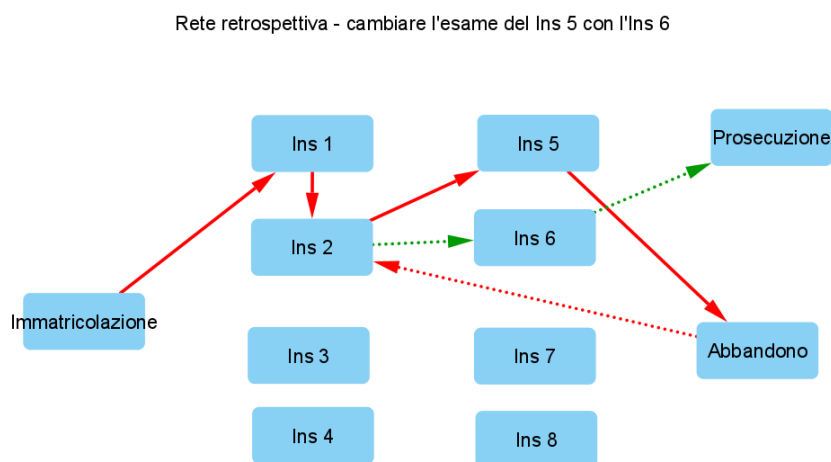
Si noti come il numero di studenti con tre esami a fine anno (7) sia minore di quelle di coloro che ne hanno superati 4 (9). Il che suggerisce di consigliare direttamente 2 ulteriori esami.

La valutazione degli effetti di una azione di orientamento è possibile solo se si **tracciano gli eventi posteriori al contatto degli studenti** a cui si è consigliato un percorso di esami. Infatti, a livello di indicatori di CdS, appare di difficile separazione l'intensità della normale variabilità temporale dall'effetto dell'intervento vero e proprio.

Inoltre, anche a livello di monitoraggio, i numeri potrebbero essere limitati poiché non tutti coloro che avrebbero bisogno di tutorato possono essere contattati e sarebbero disponibili al confronto (**tasso di adesione**) e non tutti coloro a cui si dà il consiglio poi riescono a percorrere il percorso consigliato (**tasso di conformità**). Ma non basta, per trarre delle conclusioni valide, si deve applicare l'approccio tipico del caso-controllo. Bisogna appaiare gli studenti 'consigliati' con gli studenti 'non consigliati' a parità di percorso del primo semestre e comparare l'esito finale, abbandono o prosecuzione (**tasso di successo**).

Sarebbe interessante alla fine dell'anno potere valutare percorsi del tipo descritto nella seguente figura.

Fig. 15 Rete retrospettiva



Ovvero tramite un'analisi retrospettiva, riuscire ad individuare i percorsi a rischio abbandono e si valutano percorsi aggregati "simili" che hanno portato alla prosecuzione. Simili significa, uguali nel primo semestre e diversi nel secondo, nel caso più semplice, sostituire un esame con un altro.

LE PROSPETTIVE

Si vuole, infine, accennare alla valutazione delle analisi qui proposte nella prospettiva dei principali portatori di interesse interni all'Ateneo: gli studenti, il docente, il Coordinatore del CdS, il COT.

La prospettiva dello studente

I dati mostrano una **soglia critica, 4 esami**, raggiunta la quale la probabilità di abbandono si attesta a livelli accettabili. Non solo, al di sotto della soglia fanno la differenza il semestre, **l'inattività didattica al primo semestre** spesso permane al secondo, ma, al contrario, se si supera almeno un esame il percorso **del secondo semestre è più decisivo del primo**; infine fino a tre esami, come è lecito attendersi, può fare la differenza **quali esami si superano** (a parità di numero).

La prospettiva del docente

L'abbandono dello studente **non è un evento direttamente osservabile** dal docente ne associabile al suo insegnamento. Il docente può farsi un'idea della 'stoffa' dello studente dalla qualità dell'esame o, se la numerosità del corso lo ha permesso, già durante le lezioni, ma non saprà se alla fine dell'anno quello studente avrà deciso di proseguire o meno. Se lo ha visto al primo anno raramente lo rivedrà successivamente, se non più tardi alla sessione di laurea. Se lo studente abbandona prima dell'anno di erogazione del suo insegnamento, non saprà neanche che sia esistito.

Al fini di una maggiore conoscenza del posizionamento dell'insegnamento si potrebbero periodicamente produrre le seguenti statistiche::

- quanti studenti hanno superato l'esame in ogni appello, anche in percentuale rispetto al totale. Si omettono, in questo contesto, gli esami degli iscritti ad anni successivi al primo.
- a quali altri insegnamenti si associa l'esame del proprio insegnamento
- se l'esame viene superato prima o dopo l'esame di un altro insegnamento
- percentuale di abbandono tra gli studenti che hanno superato l'esame
- percentuale di abbandono tra gli studenti che non hanno superato l'esame

Il secondo punto può essere raggiunto con la proiezione sugli insegnamenti vista in precedenza.

La prospettiva del coordinatore e del gruppo AQ

Il coordinatore, come manager del CdS, ha l'obiettivo di **pianificare e monitorare l'efficacia e l'efficienza di un'offerta formativa** che, entro i parametri AVA, sia al contempo di qualità e produttiva. Uno dei parametri AVA è la revisione annuale del CdS attraverso un commento alla Scheda di Monitoraggio Annuale (SMA).

La SMA contiene alcuni indicatori che misurano la performance al primo anno e il passaggio al secondo:

- Percentuale di CFU conseguiti al I anno su CFU da conseguire (iC13)
- Percentuale di studenti che proseguono al II anno nello stesso corso di studio (iC14), avendo acquisito almeno 20 CFU (iC15) o 40 CFU (iC16) al I anno
- Percentuale di studenti che proseguono la carriera nel sistema universitario al II anno (iC21) o in un differente CdS dell'Ateneo (iC23)

E in parte:

- Percentuale di immatricolati che si laureano entro la durata normale del Corso (iC22)
- Percentuale di abbandoni del CdS dopo N+1 anni (iC24)

Bassi valori degli indicatori sono un campanello di allarme che richiede analisi critica e azioni adeguate.

E' immediata l'utilità dell'analisi qui proposta ai fini del monitoraggio degli indicatori e l'azione di intervento mira sia ad aumentare i CFU conseguiti al primo anno sia a minimizzare la probabilità di abbandono nel passaggio al secondo anno. Inoltre lo studio qualitativo dei motivi può aiutare nella direzione di una revisione dell'offerta formativa del primo anno...come?

La prospettiva del COT

Si è mostrato come una ingegnerizzazione dei modelli di rete e di intervento possa coadiuvare il personale del COT nel momento della **consulenza di metodologia e tutorato o del counselling psicologico**, suggerendo quale possa essere il percorso del secondo semestre, alla luce del primo, che minimizzi la probabilità di abbandono.

CONCLUSIONI

Nel percorso accademico dello studente, l'anno di immatricolazione è fondamentale: un buon primo anno è indice di un buon risultato finale e viceversa un cattivo primo anno è indice o di un abbandono o comunque di un precoce rallentamento della carriera. Mantenere una carriera regolare, cercando di acquisire tutti i CFU previsti dal piano di studi è la migliore garanzia contro l'abbandono e di ottenimento della laurea nei tempi previsti. La frase è tautologica. Nessuno studente si iscrive per rimanere in ritardo. Ma può accadere e accade: le ragioni possono essere molteplici, dal difficile impatto con il sistema di studio universitario così diverso da quello della scuola superiore all'altrettanto difficile ambientamento in una nuova città nel caso dei fuori sede oppure, a volte, dall'aver scelto un Corso spinti da ragioni che si rivelano errate alla prova del primo anno universitario.

L'abbandono al primo anno è un fenomeno frequente, specialmente nelle lauree triennali e magistrali a ciclo unico, attorno al 20-25%. Le caratteristiche antecedenti il percorso del primo anno hanno un limitato potere esplicativo, tra un minimo del 10% a un massimo del 30%. La variabile più importante è il numero di esami superati durante il primo anno che discrimina tra un minimo del 2%, per coloro che superano tutti gli esami previsti dall'offerta formativa a un massimo del 90%, tra coloro che non superano alcun esame. L'inattività didattica è certamente l'evidenza più frequente di abbandono, tra tutti quelli che abbandonano la maggior parte non ha superato nessun esame. Le evidenze si colgono già al primo semestre: chi non supera almeno un esame al primo semestre, ha buone possibilità di fare altrettanto al secondo (cfr. Tab. 9). Ma anche tra coloro che superano uno o due esami al primo anno, più della metà si iscrive al secondo accumulando un ritardo che ne pregiudica la carriera successiva (cfr. Tab. 7). Si evidenzia una soglia, a livello di ateneo, di almeno 4 esami superati che si associa a tassi di abbandono accettabili, al di sotto del 10%.

L'analisi di rete ha mostrato come sia possibile associare, intuitivamente, abbandono e superamento esami, e attraverso tecniche grafiche

L'analisi dei percorsi è in grado di suggerire percorsi futuri, tipicamente esami da sostenere dopo il primo semestre, sulla base dei percorsi degli anni precedenti allo scopo di aumentare il numero di esami superati dagli studenti e minimizzare la probabilità di abbandono.

Proposte operative

L'analisi esposta in questa tesi si presta ad un processo di ingegnerizzazione e personalizzazione per CdS utile ai principali portatori di interessi nella direzione sia di una maggiore conoscenza dei meccanismi associativi tra percorso del primo anno e abbandono sia di uno strumento operativo di supporto alle decisioni e al tutoraggio in itinere.

Limitazioni

La principale limitazione ai risultati esposti risiede nel carattere esplorativo dell'analisi sviluppata che necessita di ulteriori rinforzi metodologici ed empirici. Inoltre alcuni aspetti definatori come la sovrapposizione del passaggio di CdS e la mancata iscrizione, l'esclusione di studenti con carriere pregresse o di CdS con offerte formative un po' più complesse di quelle scelte, delimitano un raggio di azione che, sebbene sia sufficientemente ampio, non può intendersi esaustivo. Inoltre le tecniche delle reti sociali sono state applicate a soli 3 CdS e necessitano dunque di una maggiore validazione. Infine, le proposte di intervento sui percorsi degli studenti si fondano su argomentazioni teoriche basate sull'assunzione di trasportabilità temporale di percorsi dagli anni pregressi all'attuale e avrebbero bisogno di una

Ricerca futura

Le future ricerche si diramano nelle distinte direzioni di un maggior rafforzamento modellistico e una conferma empirica alle assunzioni:

- formulare modelli distinti per la mancata iscrizione e il passaggio di CdS e compararne i risultati;
- prevedere l'abbandono e le variabili più influenti tramite modelli ad equazioni, ad esempio la regressione logistica, anche all'interno delle comunità estratte dalle reti;
- aggiungere altre variabili sociodemografiche, reddito, titolo di studio dei genitori, status di fuori sede, al fine di aumentare il potere predittivo dei modelli
- applicare i modelli di intervento a casi reali per validare le assunzioni teoriche

Bibliografia

- (1) Asratian, A., Denley, T., & Häggkvist, R. (1998). *Bipartite Graphs and their Applications* (Cambridge Tracts in Mathematics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511984068
- (2) Barabási, Albert-László, e Márton Pósfai. *Network science*. Cambridge University Press, 2016.
- (3) Boccaletti, S., et al. «Complex Networks: Structure and Dynamics». *Physics Reports*, vol. 424, n. 4–5, febbraio 2006, pagg. 175–308. DOI.org (Crossref), doi:10.1016/j.physrep.2005.10.009.
- (4) Csardi G, Nepusz T (2006). “The igraph software package for complex network research.” *InterJournal*, *Complex Systems*, 1695. <https://igraph.org>. Luke,
- (5) Douglas A. *A User’s Guide to Network Analysis in R*. Springer, 2015.
- (6) Kolaczyk, Eric D., Csárdi, Gábor, *Statistical Analysis of Network Data with R*. Springer, 2014.
- (7) Menczer, Filippo, et al. *A First Course in Network Science*. 2020.
- (8) Ming-Xia Li et al, *Statistically validated mobile communication networks: the evolution of motifs in European and Chinese*, 2014 *New J. Phys.* 16 083038.
- (9) R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- (10) Tumminello M, Micciche`S, Lillo F, Piilo J, Mantegna RN (2011) *Statistically Validated Networks in Bipartite Complex Systems*. *PLoS ONE* 6(3): e17994. doi:10.1371/journal.pone.0017994.
- (11) Wickham et al., (2019). *Welcome to the tidyverse*. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>