

Report Scalable and Cloud Programming

Analisi di algoritmi per la Community Detection

Giulia Giusti (Matr. 943176)
Matteo Trentin (Matr. 953888)

Indice

1 Modalità di esecuzione dei test	3
1.1 Algoritmi di Non-Overlapping Community Detection	3
1.2 Analisi di SLPA	4
2 Funzione LPA di libreria	6
2.1 Tempi di esecuzione	6
2.2 Numero community individuate	7
2.3 Calcolo della modularity	8
2.4 Calcolo della separability	9
2.5 Calcolo della density	10
3 Implementazione LPA con MapReduce	11
3.1 Tempi di esecuzione	11
3.2 Numero community individuate	12
3.3 Calcolo della modularity	13
3.4 Calcolo della separability	14
3.5 Calcolo della density	15
4 Implementazione DLPA	16
4.1 Tempi di esecuzione	16
4.2 Numero community individuate	17
4.3 Calcolo della modularity	18
4.4 Calcolo della separability	19
4.5 Calcolo della density	20
5 Implementazione LPA con Pregel	21
5.1 Tempi di esecuzione	21
5.2 Numero community individuate	22
5.3 Calcolo della modularity	23
5.4 Calcolo della separability	24
5.5 Calcolo della density	25

6	Implementazione LPA MapReduce Shuffle	26
6.1	Tempi di esecuzione	26
6.2	Numero community individuate	27
6.3	Calcolo della modularity	28
6.4	Calcolo della separability	29
6.5	Calcolo della density	30
7	Implementazione SLPA	31
7.1	Tempi di esecuzione	31
7.2	Numero community individuate	32
8	Confronto Algoritmi Non-Overlapping Community Detection	33
8.1	Confronto tempi di esecuzione	33
8.2	Confronto numero community	35
8.3	Confronto modularity	37
8.4	Confronto separability	38
8.5	Confronto density	39

1 Modalità di esecuzione dei test

1.1 Algoritmi di Non-Overlapping Community Detection

Gli algoritmi di Non-Overlapping Community Detection confrontati all'interno di questo progetto sono i seguenti:

- LPA di libreria
- LPA con MapReduce
- LPA con Pregel
- DLPA
- LPA MapReduce Shuffle

Per eseguire il test su uno di questi algoritmi è necessario specificare i seguenti argomenti:

- **--vertices <path_file_nodi_grafo>**: per specificare il file dei nodi del dataset.
- **--edges <path_file_nodi_grafo>**: per specificare il file degli archi del dataset.
- **--csv <true|false>**: per richiedere o meno il salvataggio dei risultati su un file csv.
- **--simplify <true|false>**: specificando **true** si richiede l'esecuzione dell'algoritmo sul grafo semplificato, altrimenti l'esecuzione avviene sul grafo non semplificato.
- **--metrics <true|false>**: specificando **true** si richiede il calcolo delle seguenti metriche:
 - Separability
 - Modularity
 - Density

e per la separability e la density vengono calcolate le seguenti statistiche:

- Min
 - Max
 - Media
 - Mediana
- **--algorithm <LPA|DLPA|LPA_spark|LPA_pregel|LPA_shuffle>**: per specificare l'algoritmo di Non-Overlapping Community Detection che si intende eseguire.

- **--steps <numero_step>**: per specificare il numero di step di Label Propagation.
- **--metrics <true|false>**: specificando **true** si richiede il calcolo del numero di community individuate dall'algoritmo.
- **--time <true|false>**: specificando **true** si richiede il calcolo del tempo necessario per eseguire l'algoritmo, tale tempo di calcolo viene misurato utilizzando **spark.time**.
- **--communities <true|false>**: specificando **true** si richiede il calcolo del numero di community individuate dall'algoritmo.
- **--results <path_file_output>**: per specificare il file di output.

Gli algoritmi di Non-Overlapping Community Detection sono stati confrontati in base alle seguenti caratteristiche:

- Tempo di esecuzione
- Numero di community individuate
- Media modularity
- Media separability
- Media density

al variare del numero di step.

1.2 Analisi di SLPA

Essendo SLPA l'unico algoritmo di Overlapping Community Detection implementato all'interno di questo progetto si è effettuata solamente un'analisi delle sue prestazioni senza effettuare nessun confronto.

Per eseguire il test su SLPA è necessario specificare gli argomenti nel seguente modo:

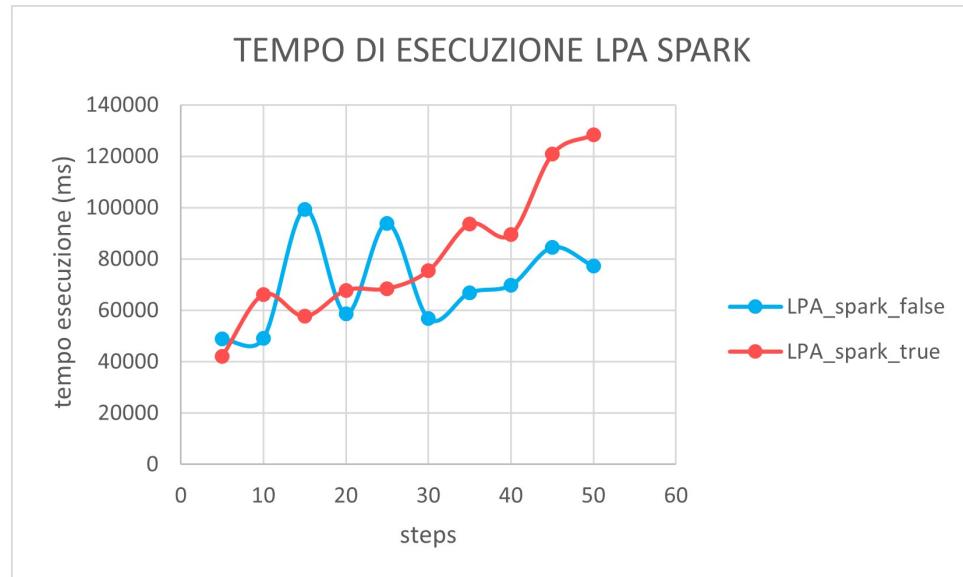
- **--vertices <path_file_nodi_grafo>**: per specificare il file dei nodi del dataset.
- **--edges <path_file_nodi_grafo>**: per specificare il file degli archi del dataset.
- **--csv <true|false>**: per richiedere o meno il salvataggio dei risultati su un file csv.
- **--metrics false**: perché le metriche elencate precedentemente non sono applicabili ai risultati degli algoritmi di Overlapping Community Detection.
- **--algorithm SLPA**

- **--steps <numero_step>**: per specificare il numero di step di Label Propagation, per SLPA il numero minimo di step è 20 e il numero massimo è 100.
- **--time <true|false>**: specificando **true** si richiede il calcolo del tempo necessario per eseguire l'algoritmo, tale tampo di calcolo viene misurato utilizzando **spark.time**.
- **--communities <true|false>**: specificando **true** si richiede il calcolo del numero di community individuate dall'algoritmo.
- **--results <path_file_output>**: per specificare il file di output.
- **--r <valore>**: soglia necessaria per la fase di post-processing dell'algoritmo, il valore di tale soglia è compreso tra [0.01, 0.1]

2 Funzione LPA di libreria

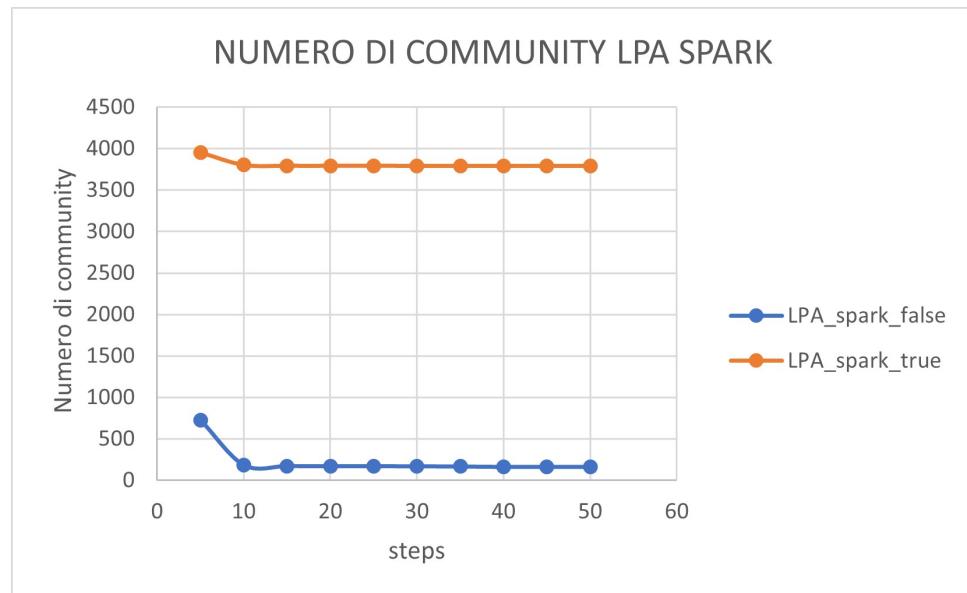
2.1 Tempi di esecuzione

N. step	Semplificato	Tempo di esecuzione
5	false	48933 ms
10	false	49189 ms
15	false	99301 ms
20	false	58737 ms
25	false	93949 ms
30	false	56843 ms
35	false	66968 ms
40	false	69851 ms
45	false	84556 ms
50	false	77196 ms
5	true	42084 ms
10	true	66195 ms
15	true	57664 ms
20	true	67866 ms
25	true	68471 ms
30	true	75458 ms
35	true	93663 ms
40	true	89475 ms
45	true	120799 ms
50	true	128393 ms



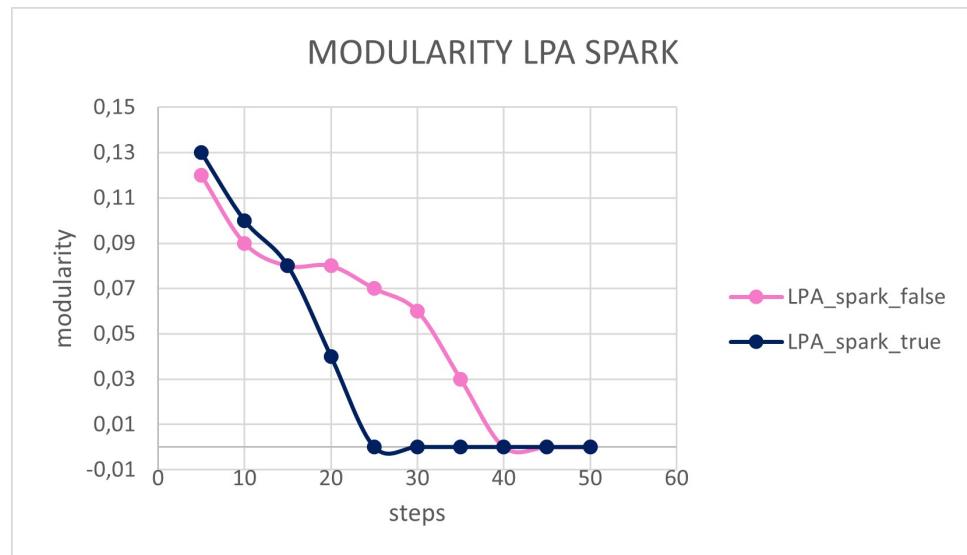
2.2 Numero community individuate

N. step	Semplificato	N. Community
5	false	726
10	false	185
15	false	173
20	false	171
25	false	171
30	false	170
35	false	167
40	false	162
45	false	162
50	false	162
5	true	3952
10	true	3802
15	true	3792
20	true	3792
25	true	3792
30	true	3791
35	true	3791
40	true	3791
45	true	3791
50	true	3791



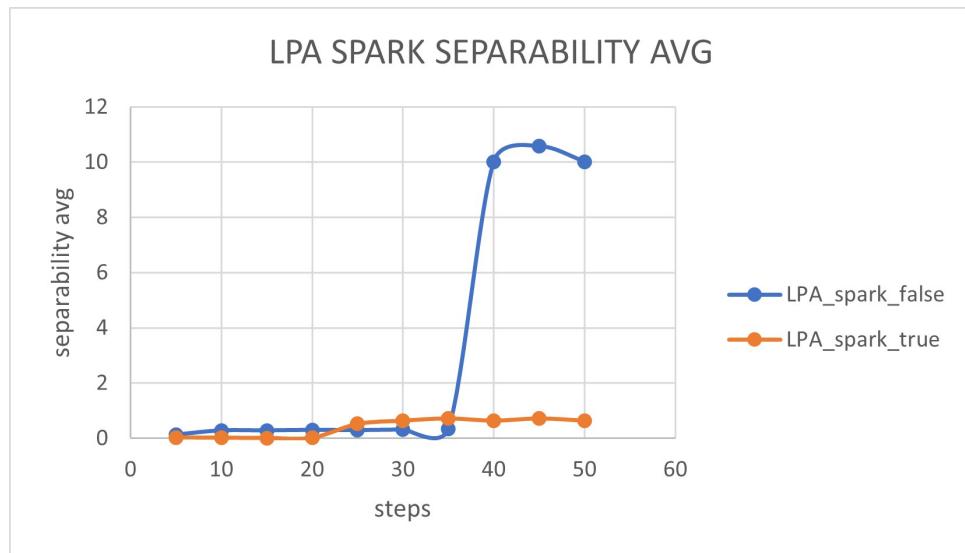
2.3 Calcolo della modularità

N. step	Semplificato	Modularity
5	false	0.12
10	false	0.09
15	false	0.08
20	false	0.08
25	false	0.07
30	false	0.06
35	false	0.03
40	false	0.0
45	false	0.0
50	false	0.0
5	true	0.13
10	true	0.1
15	true	0.08
20	true	0.04
25	true	0.0
30	true	0.0
35	true	0.0
40	true	0.0
45	true	0.0
50	true	0.0



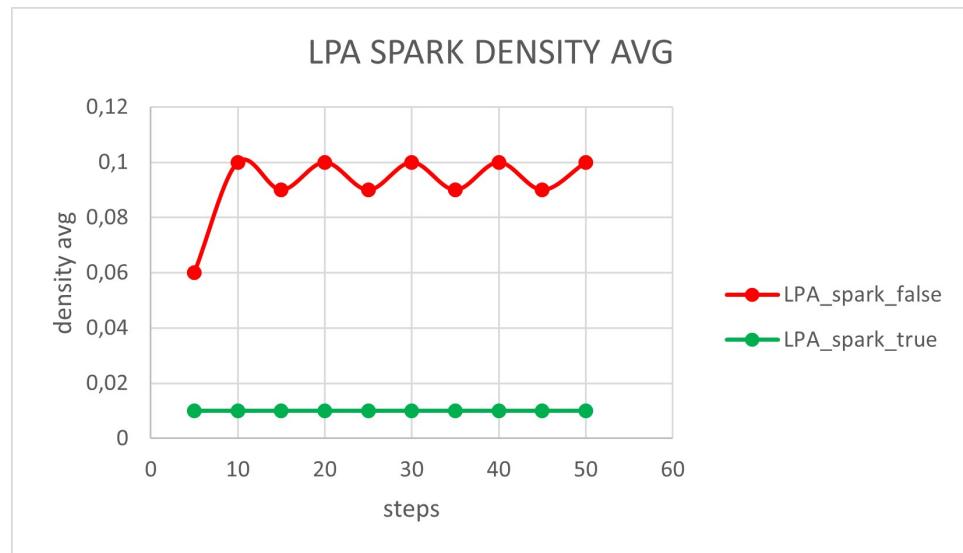
2.4 Calcolo della separability

N. step	Semplificato	Min	Max	Media	Mediana
5	false	0.0	8.7	0.13	0.0
10	false	0.0	18.41	0.28	0.0
15	false	0.0	19.23	0.28	0.0
20	false	0.0	19.62	0.3	0.0
25	false	0.0	20.48	0.29	0.0
30	false	0.0	22.5	0.31	0.0
35	false	0.0	33.63	0.35	0.0
40	false	0.0	1594.22	10.02	0.0
45	false	0.0	1687.63	10.57	0.0
50	false	0.0	1594.22	10.02	0.0
5	true	0.0	8.47	0.03	0.0
10	true	0.0	17.48	0.02	0.0
15	true	0.0	18.04	0.01	0.0
20	true	0.0	28.52	0.02	0.0
25	true	0.0	1923.62	0.52	0.0
30	true	0.0	2399.36	0.64	0.0
35	true	0.0	2699.31	0.72	0.0
40	true	0.0	2399.36	0.64	0.0
45	true	0.0	2699.31	0.72	0.0
50	true	0.0	2399.36	0.64	0.0



2.5 Calcolo della density

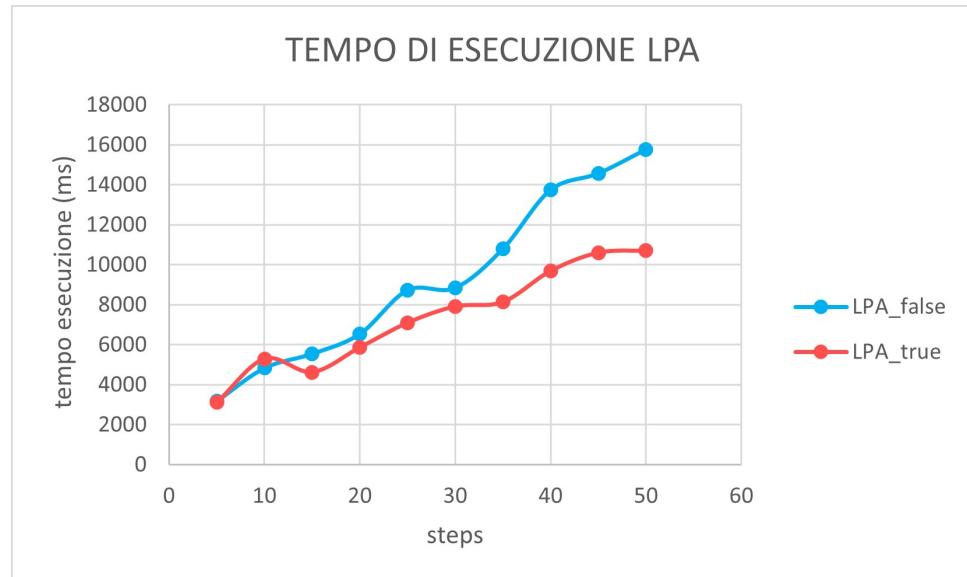
N. step	Semplificato	Min	Max	Media	Mediana
5	false	0.0	1.0	0.06	0.0
10	false	0.0	1.0	0.1	0.0
15	false	0.0	1.0	0.09	0.0
20	false	0.0	1.0	0.1	0.0
25	false	0.0	1.0	0.09	0.0
30	false	0.0	1.0	0.1	0.0
35	false	0.0	1.0	0.09	0.0
40	false	0.0	1.0	0.1	0.0
45	false	0.0	1.0	0.09	0.0
50	false	0.0	1.0	0.1	0.0
5	true	0.0	1.0	0.01	0.0
10	true	0.0	1.0	0.01	0.0
15	true	0.0	1.0	0.01	0.0
20	true	0.0	1.0	0.01	0.0
25	true	0.0	1.0	0.01	0.0
30	true	0.0	1.0	0.01	0.0
35	true	0.0	1.0	0.01	0.0
40	true	0.0	1.0	0.01	0.0
45	true	0.0	1.0	0.01	0.0
50	true	0.0	1.0	0.01	0.0



3 Implementazione LPA con MapReduce

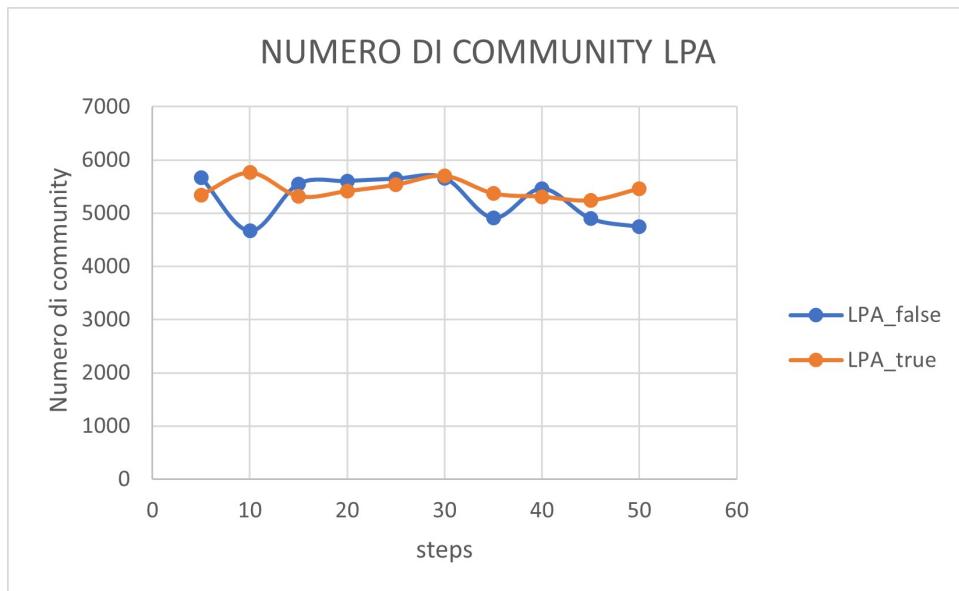
3.1 Tempi di esecuzione

N. step	Semplificato	Tempo di esecuzione
5	false	3186 ms
10	false	4839 ms
15	false	5552 ms
20	false	6548 ms
25	false	8735 ms
30	false	8849 ms
35	false	10803 ms
40	false	13747 ms
45	false	14575 ms
50	false	15773 ms
5	true	3112 ms
10	true	5292 ms
15	true	4627 ms
20	true	5862 ms
25	true	7095 ms
30	true	7915 ms
35	true	8136 ms
40	true	9688 ms
45	true	10600 ms
50	true	10714 ms



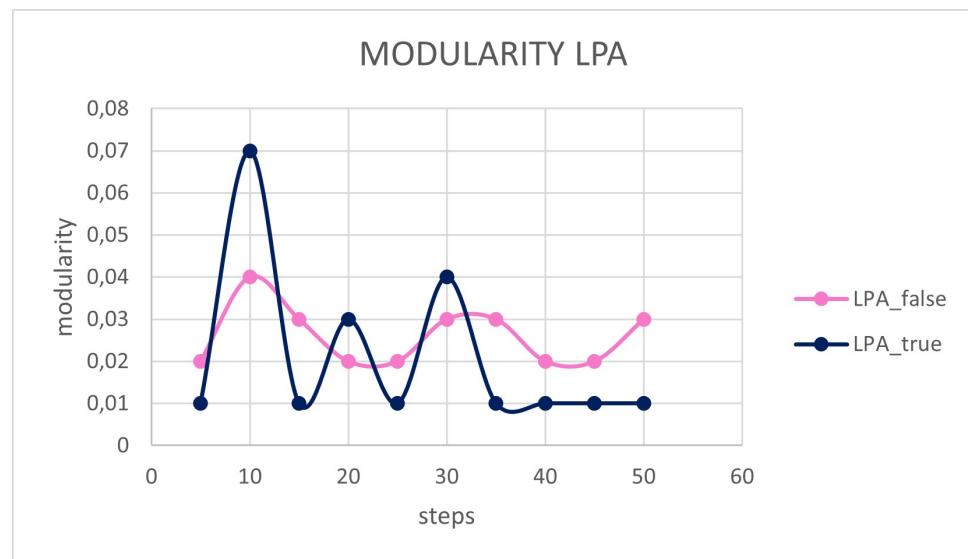
3.2 Numero community individuate

N. step	Semplificato	N. Community
5	false	5664
10	false	4667
15	false	5548
20	false	5601
25	false	5646
30	false	5661
35	false	4911
40	false	5460
45	false	4899
50	false	4743
5	true	5344
10	true	5763
15	true	5316
20	true	5414
25	true	5534
30	true	5703
35	true	5371
40	true	5309
45	true	5240
50	true	5459



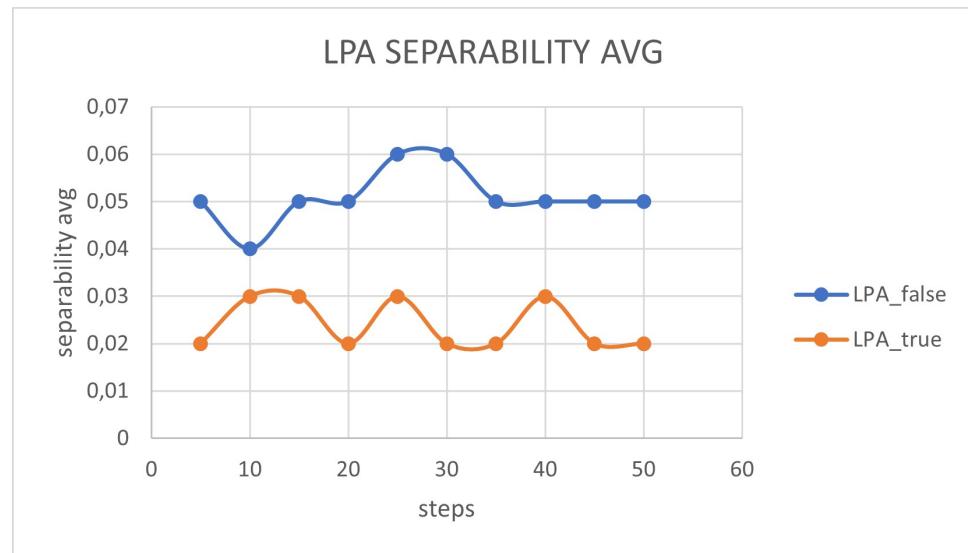
3.3 Calcolo della modularità

N. step	Semplificato	Modularity
5	false	0.02
10	false	0.04
15	false	0.03
20	false	0.02
25	false	0.02
30	false	0.03
35	false	0.03
40	false	0.02
45	false	0.02
50	false	0.03
5	true	0.01
10	true	0.07
15	true	0.01
20	true	0.03
25	true	0.01
30	true	0.04
35	true	0.01
40	true	0.01
45	true	0.01
50	true	0.01



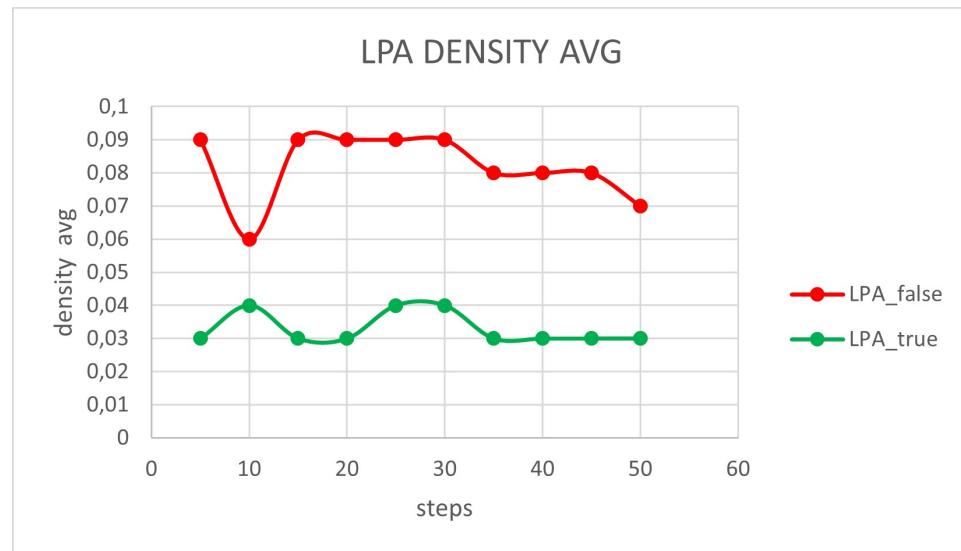
3.4 Calcolo della separability

N. step	Semplificato	Min	Max	Media	Mediana
5	false	0.0	4.0	0.05	0.0
10	false	0.0	4.0	0.04	0.0
15	false	0.0	4.0	0.05	0.0
20	false	0.0	9.13	0.05	0.0
25	false	0.0	4.0	0.06	0.0
30	false	0.0	7.88	0.06	0.0
35	false	0.0	3.0	0.05	0.0
40	false	0.0	14.74	0.05	0.0
45	false	0.0	4.0	0.05	0.0
50	false	0.0	4.0	0.05	0.0
5	true	0.0	5.0	0.02	0.0
10	true	0.0	8.05	0.03	0.0
15	true	0.0	43.45	0.03	0.0
20	true	0.0	1.8	0.02	0.0
25	true	0.0	30.45	0.03	0.0
30	true	0.0	1.8	0.02	0.0
35	true	0.0	30.79	0.02	0.0
40	true	0.0	43.07	0.03	0.0
45	true	0.0	4.67	0.02	0.0
50	true	0.0	4.0	0.02	0.0



3.5 Calcolo della density

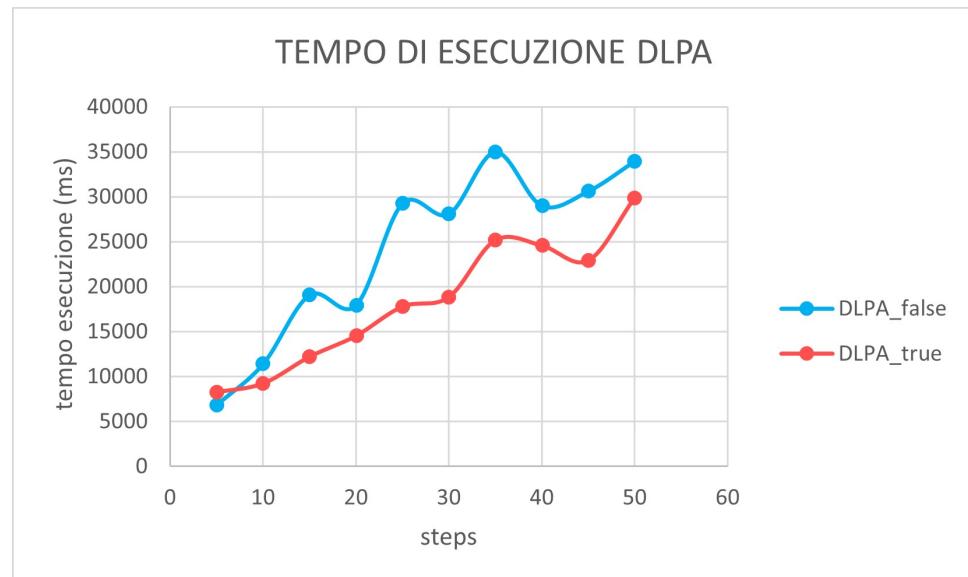
N. step	Semplificato	Min	Max	Media	Median
5	false	0.0	1.0	0.09	0.0
10	false	0.0	1.0	0.06	0.0
15	false	0.0	1.0	0.09	0.0
20	false	0.0	1.0	0.09	0.0
25	false	0.0	1.0	0.09	0.0
30	false	0.0	1.0	0.09	0.0
35	false	0.0	1.0	0.08	0.0
40	false	0.0	1.0	0.08	0.0
45	false	0.0	1.0	0.08	0.0
50	false	0.0	1.0	0.07	0.0
5	true	0.0	1.0	0.03	0.0
10	true	0.0	1.0	0.04	0.0
15	true	0.0	1.0	0.03	0.0
20	true	0.0	1.0	0.03	0.0
25	true	0.0	1.0	0.04	0.0
30	true	0.0	1.0	0.04	0.0
35	true	0.0	1.0	0.03	0.0
40	true	0.0	1.0	0.03	0.0
45	true	0.0	1.0	0.03	0.0
50	true	0.0	1.0	0.03	0.0



4 Implementazione DLPA

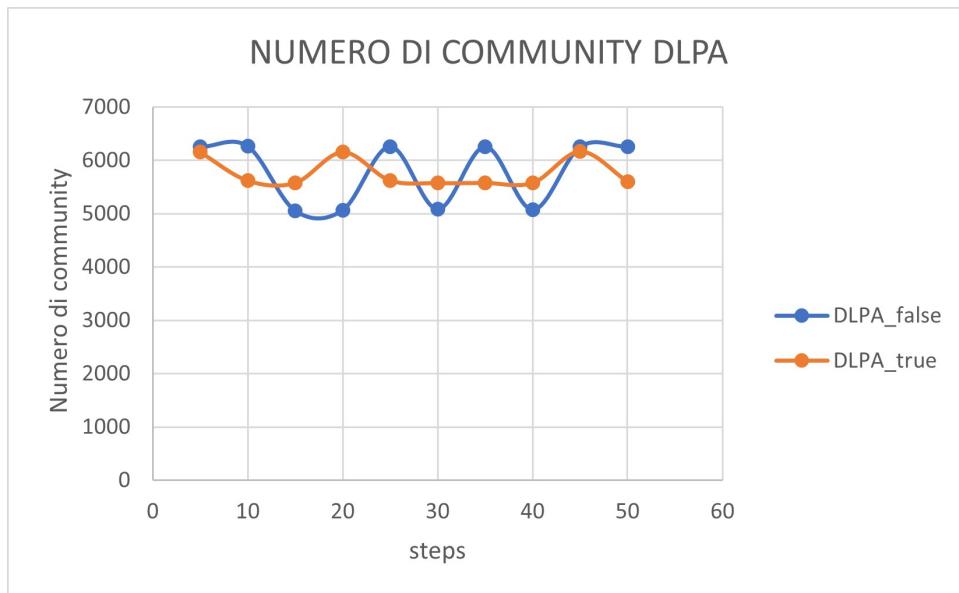
4.1 Tempi di esecuzione

N. step	Semplificato	Tempo di esecuzione
5	false	6844 ms
10	false	11416 ms
15	false	19085 ms
20	false	17944 ms
25	false	29273 ms
30	false	28104 ms
35	false	34995 ms
40	false	29030 ms
45	false	30624 ms
50	false	33955 ms
5	true	8263 ms
10	true	9245 ms
15	true	12218 ms
20	true	14539 ms
25	true	17801 ms
30	true	18859 ms
35	true	25185 ms
40	true	24600 ms
45	true	22916 ms
50	true	29879 ms



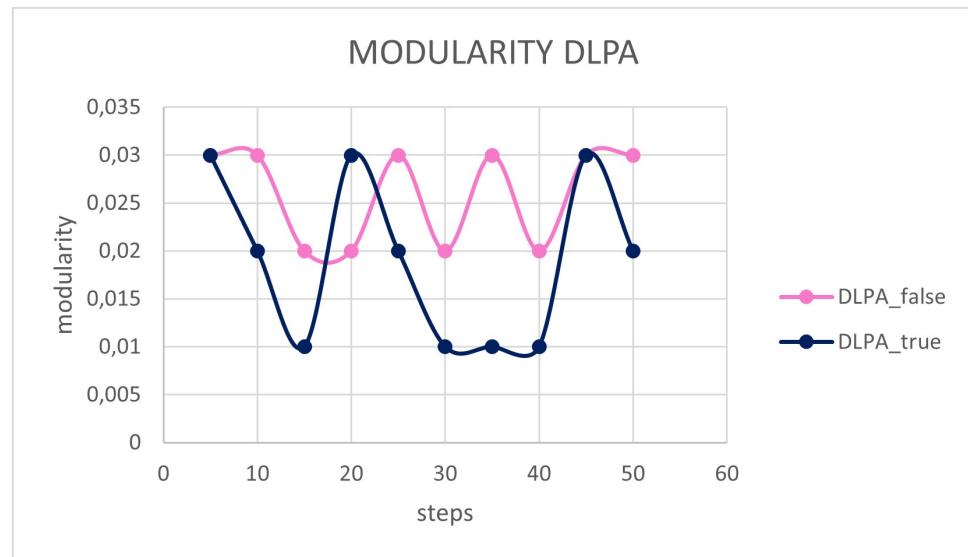
4.2 Numero community individuate

N. step	Semplificato	N. Community
5	false	6252
10	false	6267
15	false	5056
20	false	5067
25	false	6260
30	false	5087
35	false	6257
40	false	5075
45	false	6258
50	false	6255
5	true	6158
10	true	5622
15	true	5580
20	true	6158
25	true	5624
30	true	5577
35	true	5580
40	true	5580
45	true	6169
50	true	5602



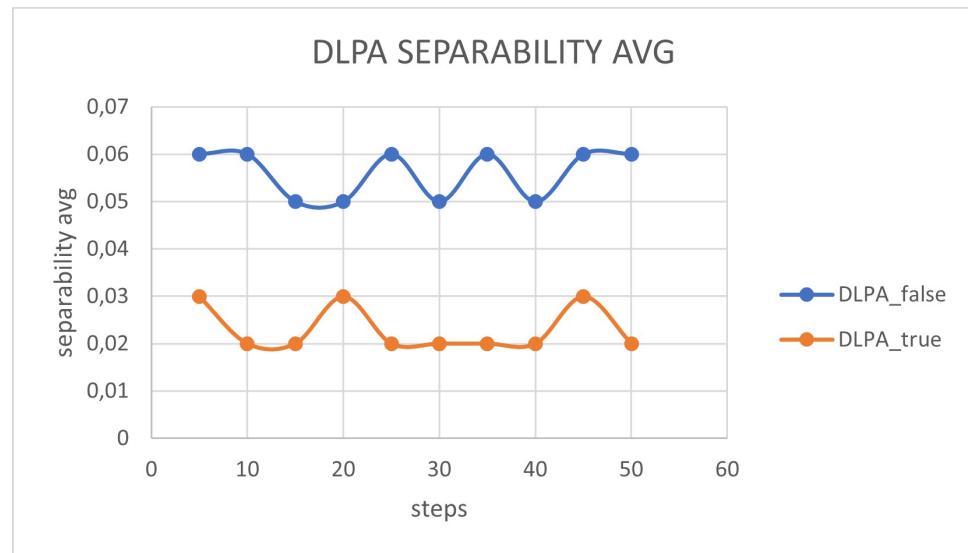
4.3 Calcolo della modularity

N. step	Semplificato	Modularity
5	false	0.03
10	false	0.03
15	false	0.02
20	false	0.02
25	false	0.03
30	false	0.02
35	false	0.03
40	false	0.02
45	false	0.03
50	false	0.03
5	true	0.03
10	true	0.02
15	true	0.01
20	true	0.03
25	true	0.02
30	true	0.01
35	true	0.01
40	true	0.01
45	true	0.03
50	true	0.02



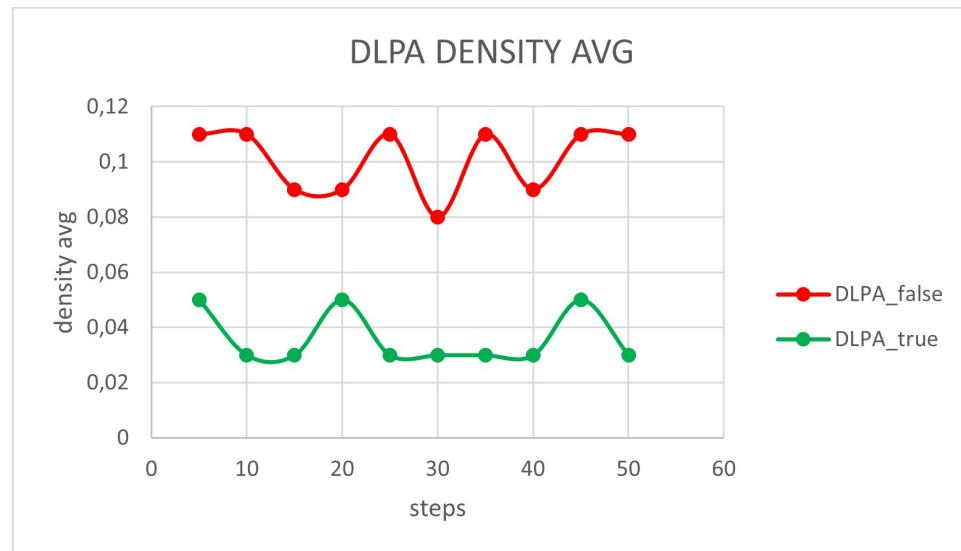
4.4 Calcolo della separability

N. step	Semplificato	Min	Max	Media	Mediana
5	false	0.0	4.0	0.06	0.0
10	false	0.0	4.0	0.06	0.0
15	false	0.0	4.0	0.05	0.0
20	false	0.0	4.0	0.05	0.0
25	false	0.0	4.0	0.06	0.0
30	false	0.0	4.0	0.05	0.0
35	false	0.0	4.0	0.06	0.0
40	false	0.0	4.0	0.05	0.0
45	false	0.0	4.0	0.06	0.0
50	false	0.0	4.0	0.06	0.0
5	true	0.0	4.0	0.03	0.0
10	true	0.0	3.0	0.02	0.0
15	true	0.0	5.0	0.02	0.0
20	true	0.0	4.5	0.03	0.0
25	true	0.0	4.5	0.02	0.0
30	true	0.0	5.0	0.02	0.0
35	true	0.0	5.0	0.02	0.0
40	true	0.0	5.0	0.02	0.0
45	true	0.0	4.5	0.03	0.0
50	true	0.0	5.0	0.02	0.0



4.5 Calcolo della density

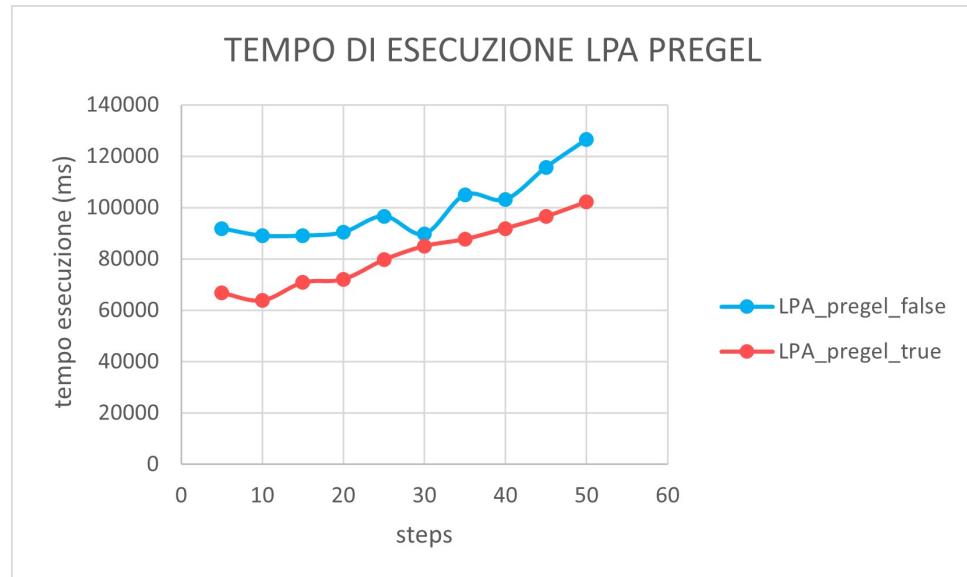
N. step	Semplificato	Min	Max	Media	Median
5	false	0.0	1.0	0.11	0.0
10	false	0.0	1.0	0.11	0.0
15	false	0.0	1.0	0.09	0.0
20	false	0.0	1.0	0.09	0.0
25	false	0.0	1.0	0.11	0.0
30	false	0.0	1.0	0.08	0.0
35	false	0.0	1.0	0.11	0.0
40	false	0.0	1.0	0.09	0.0
45	false	0.0	1.0	0.11	0.0
50	false	0.0	1.0	0.11	0.0
5	true	0.0	1.0	0.05	0.0
10	true	0.0	1.0	0.03	0.0
15	true	0.0	1.0	0.03	0.0
20	true	0.0	1.0	0.05	0.0
25	true	0.0	1.0	0.03	0.0
30	true	0.0	1.0	0.03	0.0
35	true	0.0	1.0	0.03	0.0
40	true	0.0	1.0	0.03	0.0
45	true	0.0	1.0	0.05	0.0
50	true	0.0	1.0	0.03	0.0



5 Implementazione LPA con Pregel

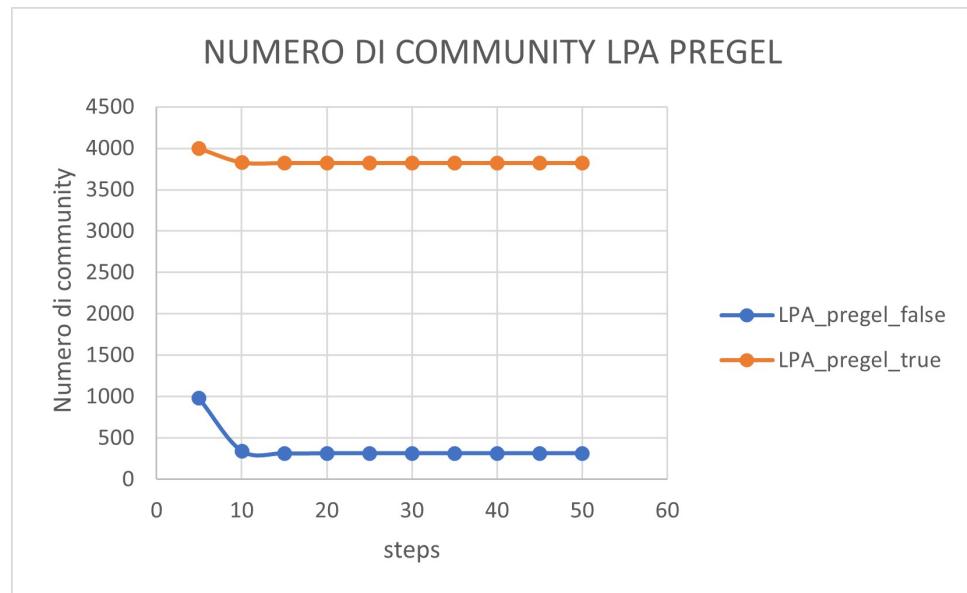
5.1 Tempi di esecuzione

N. step	Semplificato	Tempo di esecuzione
5	false	91838 ms
10	false	89012 ms
15	false	89026 ms
20	false	90320 ms
25	false	96620 ms
30	false	89683 ms
35	false	104933 ms
40	false	103111 ms
45	false	115611 ms
50	false	126445 ms
5	true	66840 ms
10	true	63888 ms
15	true	70832 ms
20	true	72114 ms
25	true	79646 ms
30	true	85000 ms
35	true	87681 ms
40	true	91872 ms
45	true	96479 ms
50	true	102139 ms



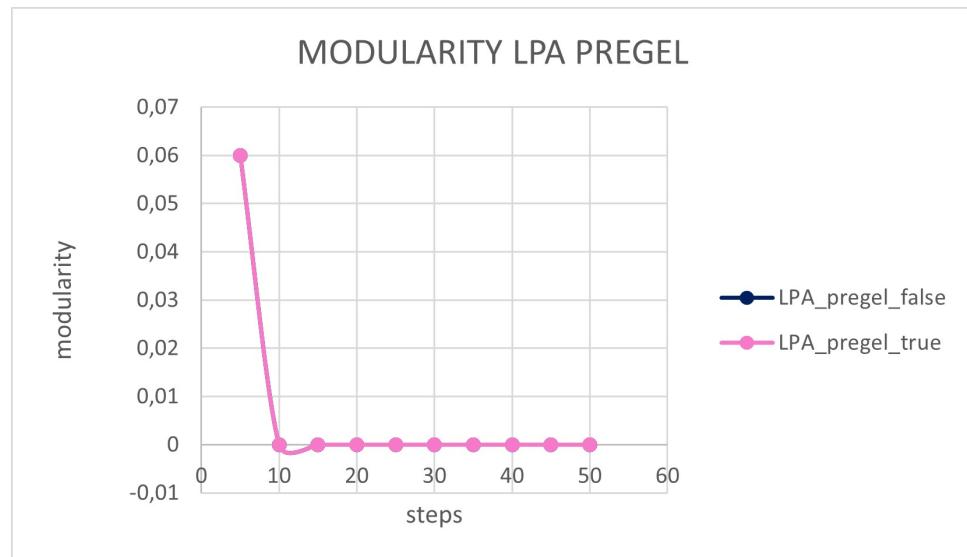
5.2 Numero community individuate

N. step	Semplificato	N. Community
5	false	980
10	false	343
15	false	313
20	false	313
25	false	313
30	false	313
35	false	313
40	false	313
45	false	313
50	false	313
5	true	4000
10	true	3830
15	true	3823
20	true	3823
25	true	3823
30	true	3823
35	true	3823
40	true	3823
45	true	3823
50	true	3823



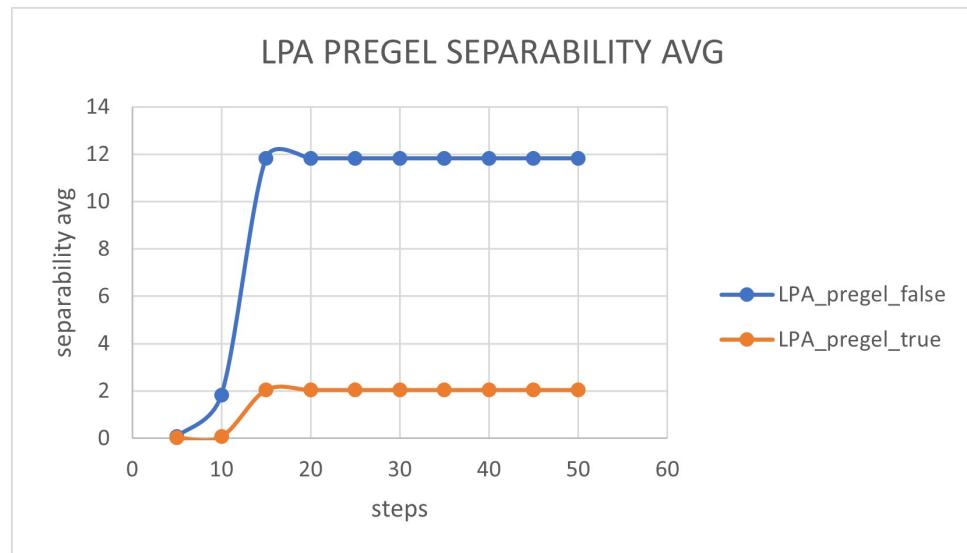
5.3 Calcolo della modularità

N. step	Semplificato	Modularity
5	false	0.06
10	false	0.0
15	false	0.0
20	false	0.0
25	false	0.0
30	false	0.0
35	false	0.0
40	false	0.0
45	false	0.0
50	false	0.0
5	true	0.06
10	true	0.0
15	true	0.0
20	true	0.0
25	true	0.0
30	true	0.0
35	true	0.0
40	true	0.0
45	true	0.0
50	true	0.0



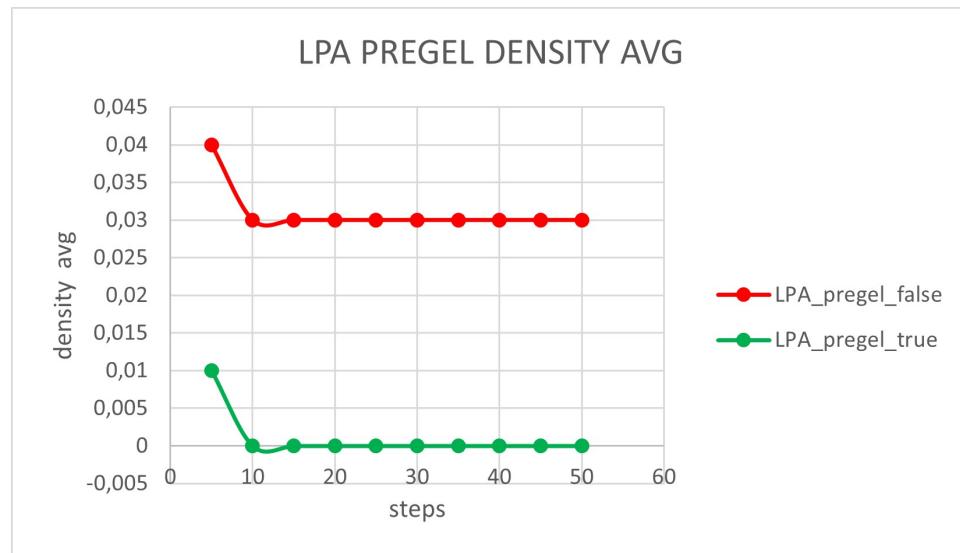
5.4 Calcolo della separability

N. step	Semplificato	Min	Max	Media	Mediana
5	false	0.0	7.29	0.07	0.0
10	false	0.0	622.6	1.83	0.0
15	false	0.0	3699.24	11.83	0.0
20	false	0.0	3699.24	11.83	0.0
25	false	0.0	3699.24	11.83	0.0
30	false	0.0	3699.24	11.83	0.0
35	false	0.0	3699.24	11.83	0.0
40	false	0.0	3699.24	11.83	0.0
45	false	0.0	3699.24	11.83	0.0
50	false	0.0	3699.24	11.83	0.0
5	true	0.0	8.89	0.02	0.0
10	true	0.0	269.32	0.08	0.0
15	true	0.0	7775.72	2.04	0.0
20	true	0.0	7775.72	2.04	0.0
25	true	0.0	7775.72	2.04	0.0
30	true	0.0	7775.72	2.04	0.0
35	true	0.0	7775.72	2.04	0.0
40	true	0.0	7775.72	2.04	0.0
45	true	0.0	7775.72	2.04	0.0
50	true	0.0	7775.72	2.04	0.0



5.5 Calcolo della density

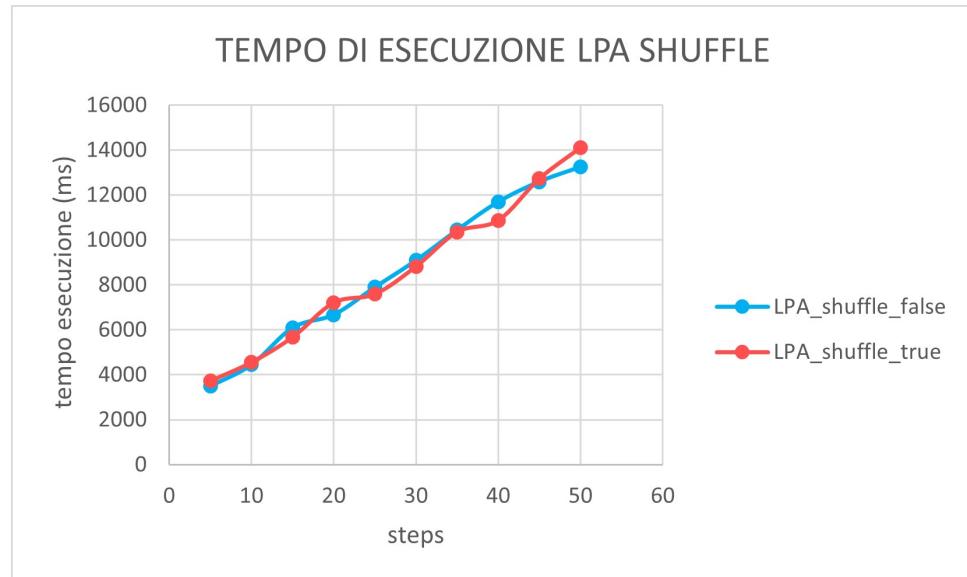
N. step	Semplificato	Min	Max	Media	Mediana
5	false	0.0	1.0	0.04	0.0
10	false	0.0	1.0	0.03	0.0
15	false	0.0	1.0	0.03	0.0
20	false	0.0	1.0	0.03	0.0
25	false	0.0	1.0	0.03	0.0
30	false	0.0	1.0	0.03	0.0
35	false	0.0	1.0	0.03	0.0
40	false	0.0	1.0	0.03	0.0
45	false	0.0	1.0	0.03	0.0
50	false	0.0	1.0	0.03	0.0
5	true	0.0	1.0	0.01	0.0
10	true	0.0	1.0	0.0	0.0
15	true	0.0	1.0	0.0	0.0
20	true	0.0	1.0	0.0	0.0
25	true	0.0	1.0	0.0	0.0
30	true	0.0	1.0	0.0	0.0
35	true	0.0	1.0	0.0	0.0
40	true	0.0	1.0	0.0	0.0
45	true	0.0	1.0	0.0	0.0
50	true	0.0	1.0	0.0	0.0



6 Implementazione LPA MapReduce Shuffle

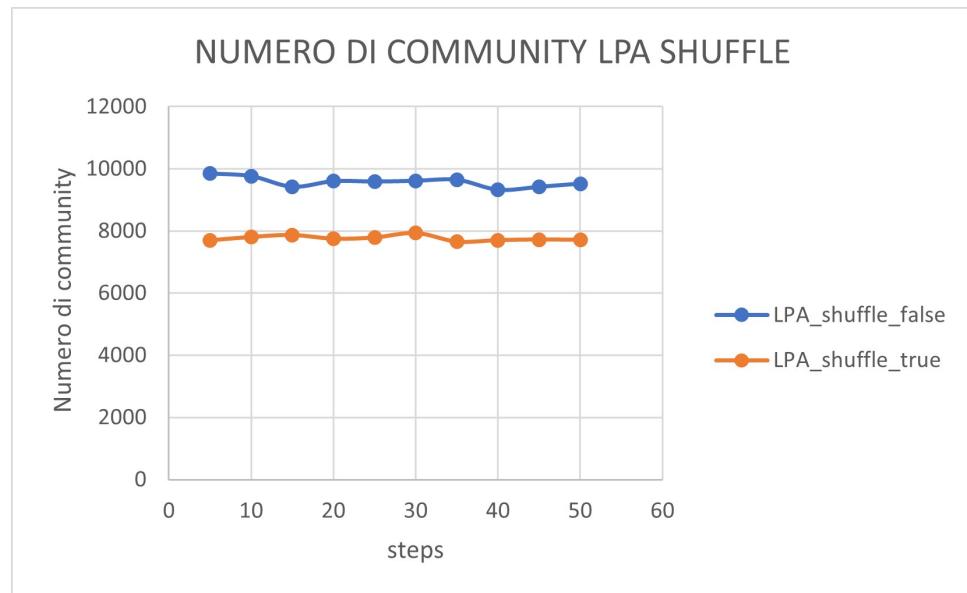
6.1 Tempi di esecuzione

N. step	Semplificato	Tempo di esecuzione
5	false	3490 ms
10	false	4449 ms
15	false	6095 ms
20	false	6672 ms
25	false	7897 ms
30	false	9090 ms
35	false	10440 ms
40	false	11694 ms
45	false	12593 ms
50	false	13245 ms
5	true	3729 ms
10	true	4559 ms
15	true	5682 ms
20	true	7214 ms
25	true	7593 ms
30	true	8827 ms
35	true	10360 ms
40	true	10876 ms
45	true	12732 ms
50	true	14111 ms



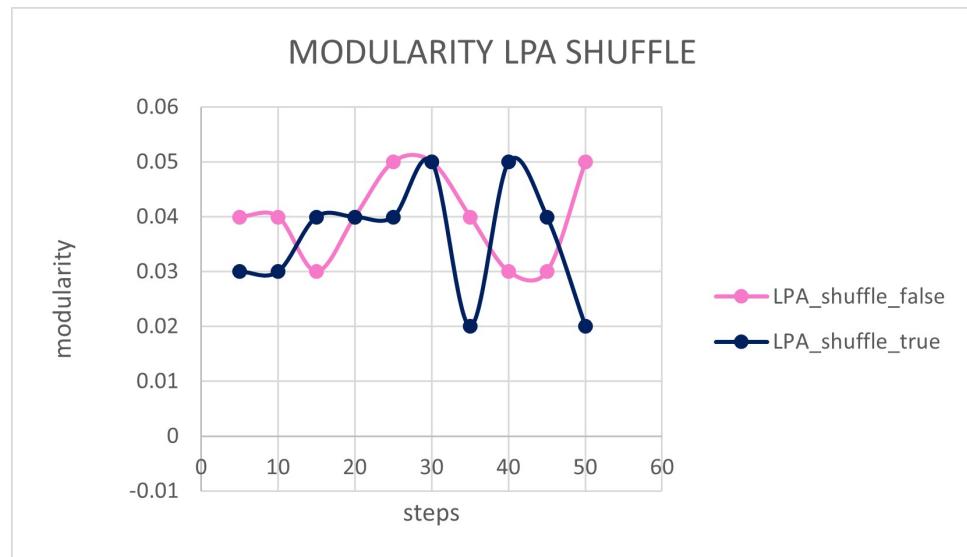
6.2 Numero community individuate

N. step	Semplificato	N. Community
5	false	9850
10	false	9764
15	false	9422
20	false	9602
25	false	9589
30	false	9614
35	false	9647
40	false	9324
45	false	9420
50	false	9524
5	true	7693
10	true	7806
15	true	7870
20	true	7749
25	true	7789
30	true	7935
35	true	7653
40	true	7702
45	true	7724
50	true	7713



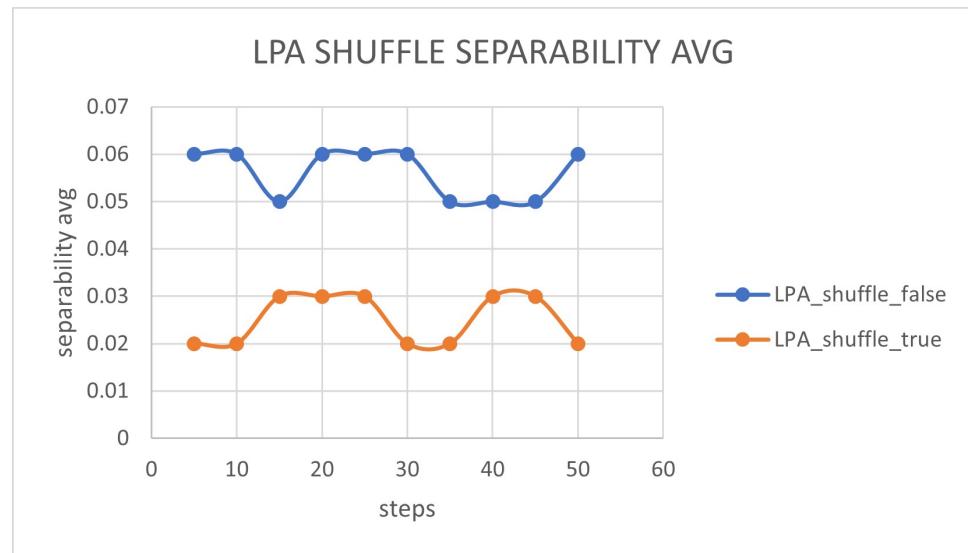
6.3 Calcolo della modularità

N. step	Semplificato	Modularity
5	false	0.04
10	false	0.04
15	false	0.03
20	false	0.04
25	false	0.05
30	false	0.05
35	false	0.04
40	false	0.03
45	false	0.03
50	false	0.05
5	true	0.03
10	true	0.03
15	true	0.04
20	true	0.04
25	true	0.04
30	true	0.05
35	true	0.02
40	true	0.05
45	true	0.04
50	true	0.02



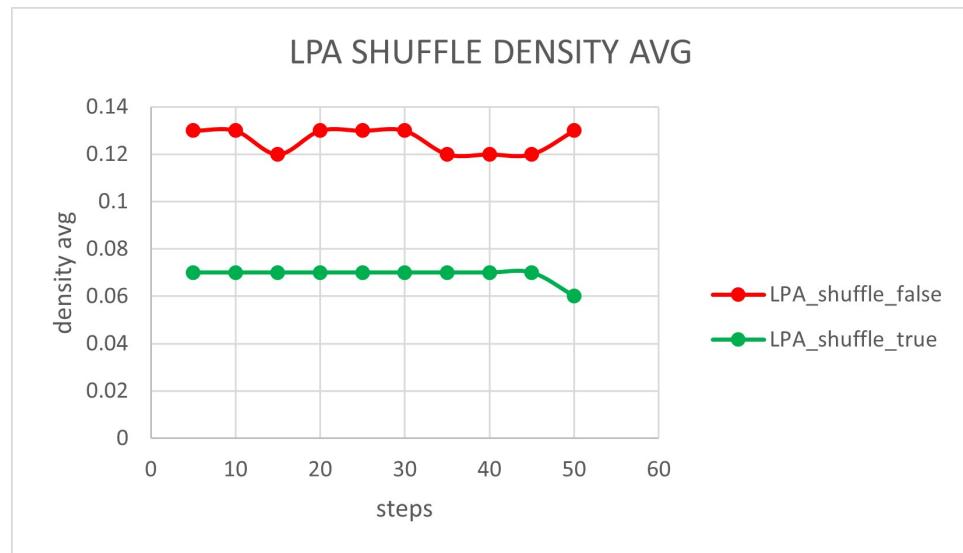
6.4 Calcolo della separability

N. step	Semplificato	Min	Max	Media	Mediana
5	false	0.0	4.0	0.06	0.0
10	false	0.0	4.0	0.06	0.0
15	false	0.0	3.0	0.05	0.0
20	false	0.0	6.0	0.06	0.0
25	false	0.0	3.33	0.06	0.0
30	false	0.0	4.0	0.06	0.0
35	false	0.0	3.0	0.05	0.0
40	false	0.0	4.0	0.05	0.0
45	false	0.0	4.0	0.05	0.0
50	false	0.0	3.0	0.06	0.0
5	true	0.0	2.5	0.02	0.0
10	true	0.0	3.0	0.02	0.0
15	true	0.0	3.0	0.03	0.0
20	true	0.0	5.0	0.03	0.0
25	true	0.0	3.5	0.03	0.0
30	true	0.0	3.0	0.02	0.0
35	true	0.0	5.0	0.02	0.0
40	true	0.0	3.0	0.03	0.0
45	true	0.0	4.0	0.03	0.0
50	true	0.0	3.0	0.02	0.0



6.5 Calcolo della density

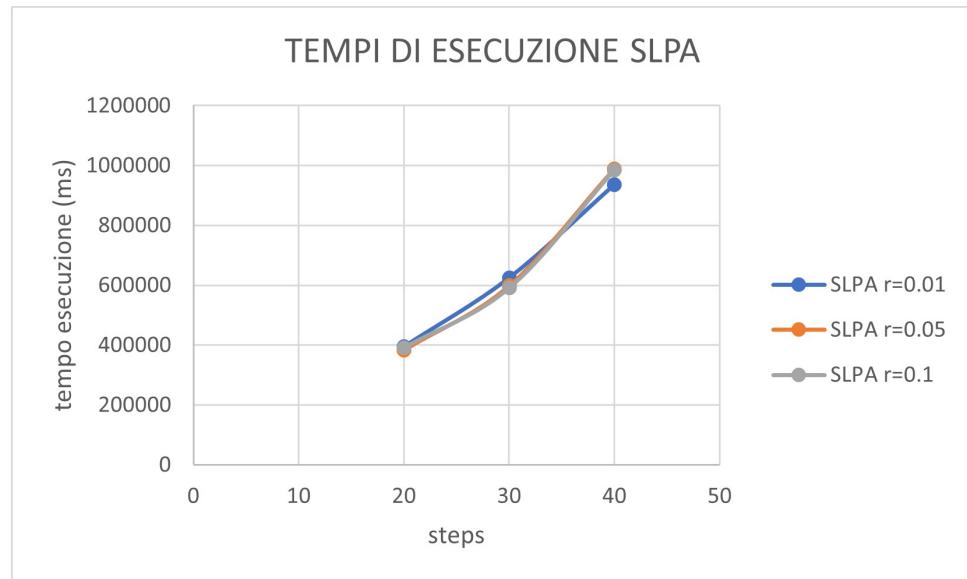
N. step	Semplificato	Min	Max	Media	Mediana
5	false	0.0	1.0	0.13	0.0
10	false	0.0	1.0	0.13	0.0
15	false	0.0	1.0	0.12	0.0
20	false	0.0	1.0	0.13	0.0
25	false	0.0	1.0	0.13	0.0
30	false	0.0	1.0	0.13	0.0
35	false	0.0	1.0	0.12	0.0
40	false	0.0	1.0	0.12	0.0
45	false	0.0	1.0	0.12	0.0
50	false	0.0	1.0	0.13	0.0
5	true	0.0	1.0	0.07	0.0
10	true	0.0	1.0	0.07	0.0
15	true	0.0	1.0	0.07	0.0
20	true	0.0	1.0	0.07	0.0
25	true	0.0	1.0	0.07	0.0
30	true	0.0	1.0	0.07	0.0
35	true	0.0	1.0	0.07	0.0
40	true	0.0	1.0	0.07	0.0
45	true	0.0	1.0	0.07	0.0
50	true	0.0	1.0	0.06	0.0



7 Implementazione SLPA

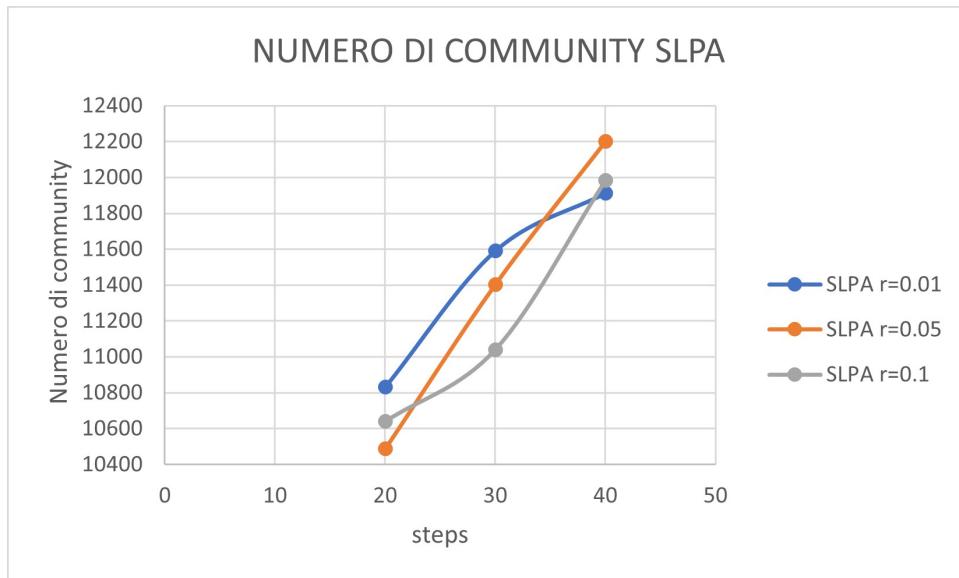
7.1 Tempi di esecuzione

N. step	r	N. Community
20	0.01	394257 ms
30	0.01	624931 ms
40	0.01	937003 ms
20	0.05	384287 ms
30	0.05	598557 ms
40	0.05	988928 ms
20	0.1	391468 ms
30	0.1	591545 ms
40	0.1	985528 ms



7.2 Numero community individuate

N. step	r	N. Community
20	0.01	10831
30	0.01	11591
40	0.01	11913
20	0.05	10487
30	0.05	11404
40	0.05	12201
20	0.1	10639
30	0.1	11041
40	0.1	11984

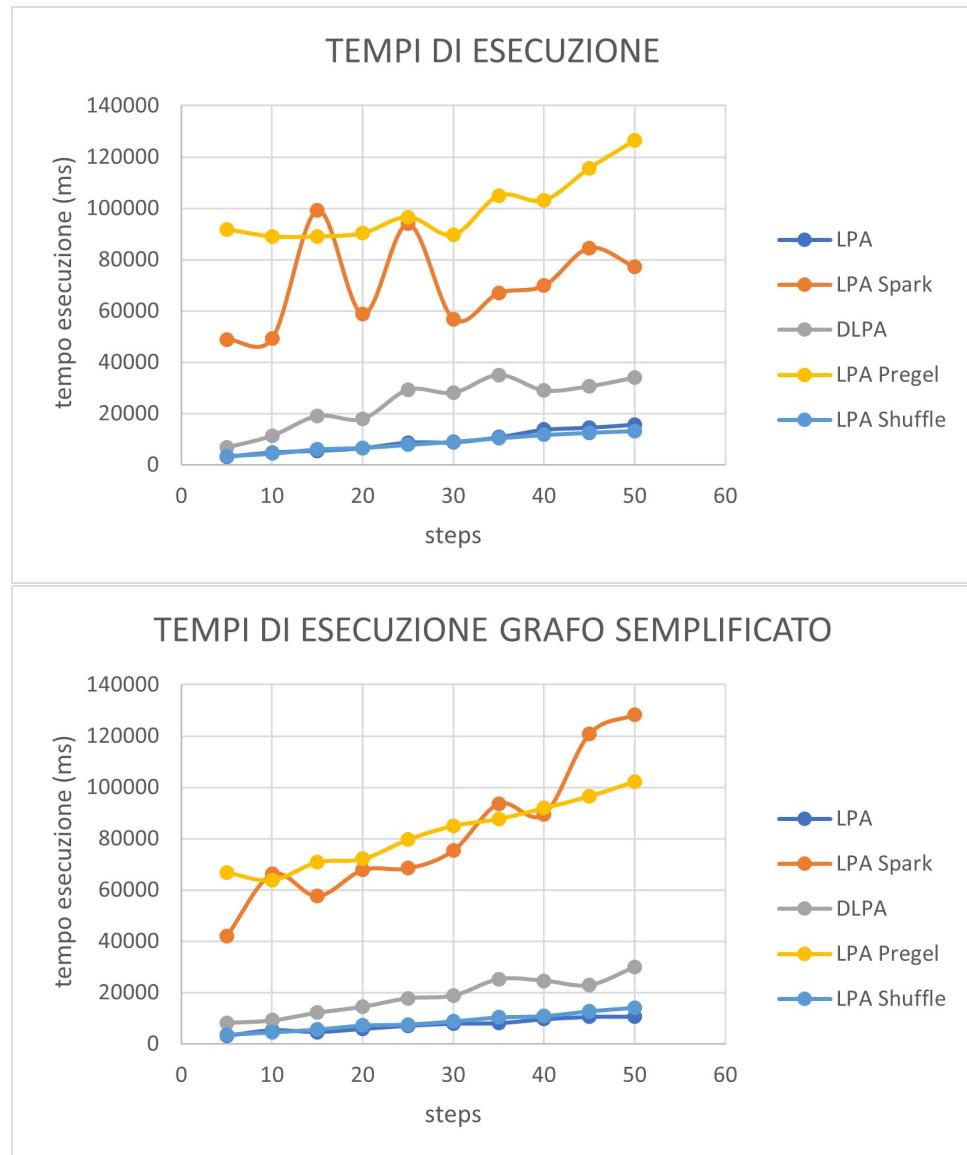


8 Confronto Algoritmi Non-Overlapping Community Detection

8.1 Confronto tempi di esecuzione

I risultati sono stati ottenuti su:

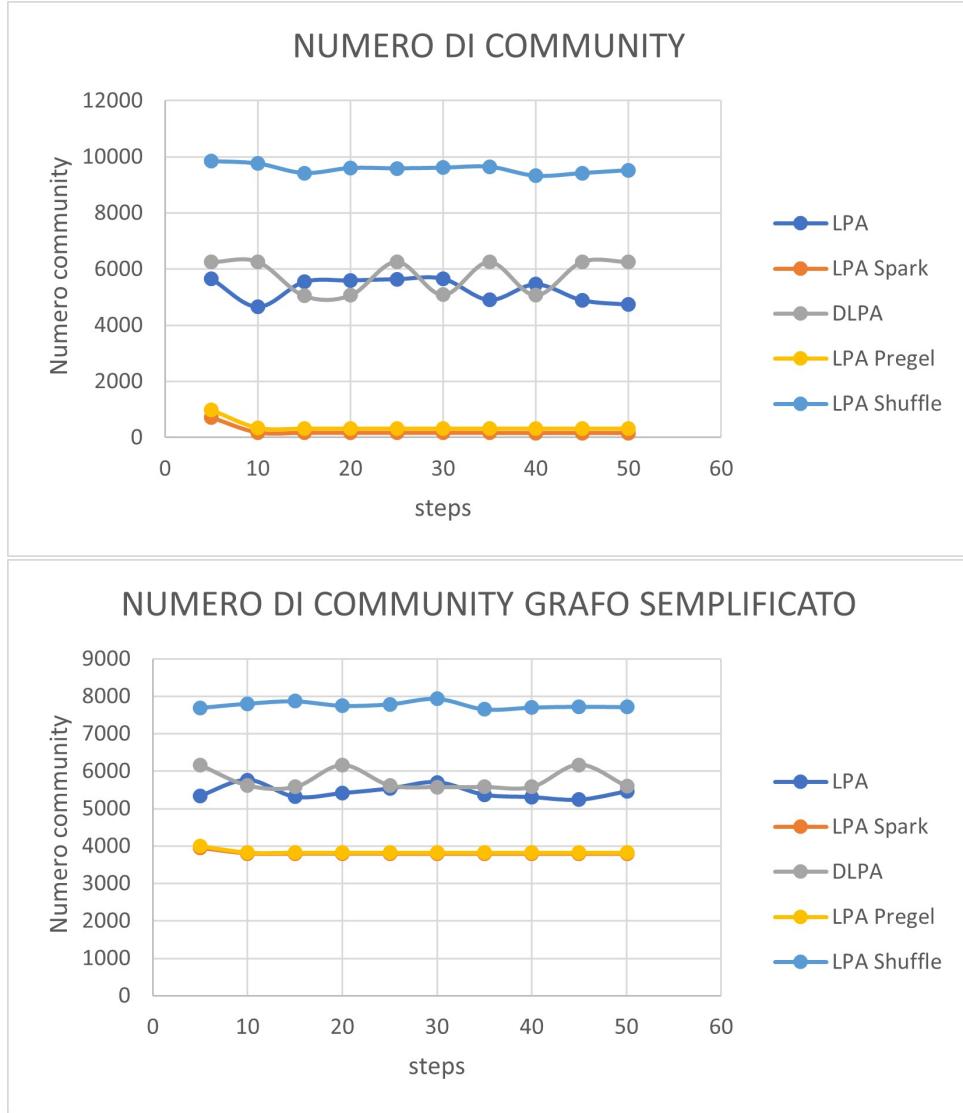
- Intel Core i7-6500U 2.50 GHz
- RAM 12.0 GB



I risultati relativi ai tempi di esecuzione sono molto simili a quanto ci aspettavamo, in quanto:

- LPA e LPA Shuffle hanno tempi di esecuzione molto vicini perché sono entrambi basati sulla medesima struttura di Label Propagation con MapReduce e l'unica differenza tra le loro implementazioni è rappresentata dall'ordine con cui si scorrono i nodi del grafo nella fase di propagazione delle label.
- DLPA ha tempi di esecuzione leggermente più alti rispetto a LPA e LPA Shuffle poiché, essendo anch'esso basato sulla medesima struttura di Label Propagation con MapReduce, la richiesta aggiuntiva del conteggio di: inDegree, outDegree e degree di ogni nodo del grafo per il calcolo del peso del nodo richiede maggiore tempo di esecuzione.
- LPA Spark e LPA Pregel richiedono entrambi maggior tempo di esecuzione rispetto ai precedenti perché sono basata su una differente struttura di Label Propagation basata sul framework Pregel.

8.2 Confronto numero community



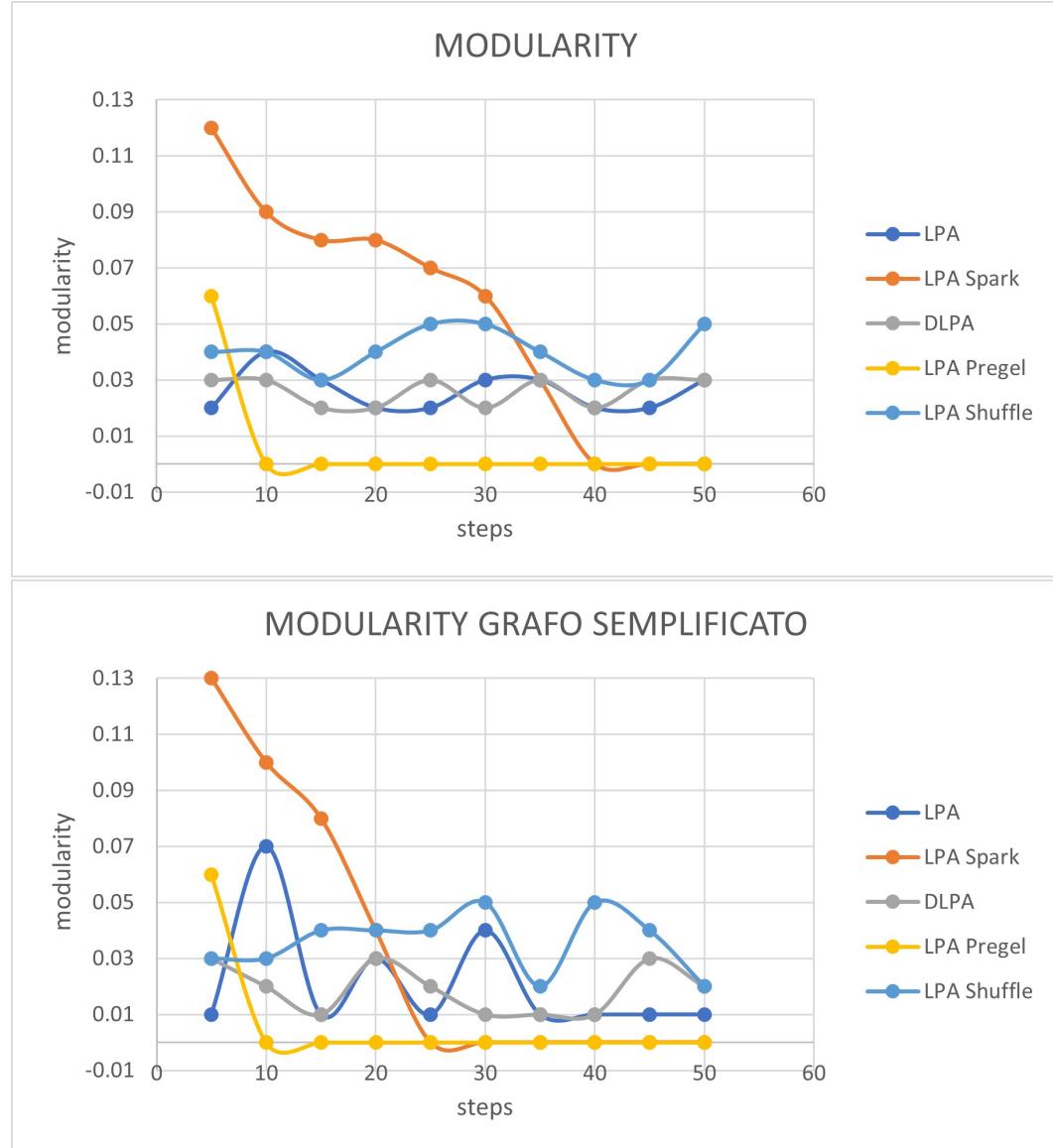
I risultati relativi al numero di community rilevate dai diversi algoritmi sono, in alcuni casi, abbastanza sorprendenti:

- LPA Spark e LPA Pregel essendo basati sullo stesso framework Pregel ed essendo le loro implementazioni molto simili restituiscono risultati molto simili in termini di numero di community individuate.
- LPA Shuffle grazie all'estrazione casuale dei nodi all'interno dell'insieme dei vertici del grafo durante la fase di Label Propagation dovrebbe aiutare l'algoritmo ad identificare un numero minore di community e ognuna di

esse dovrebbe essere di maggiore qualità, questo però non viene riscontrato nei nostri risultati.

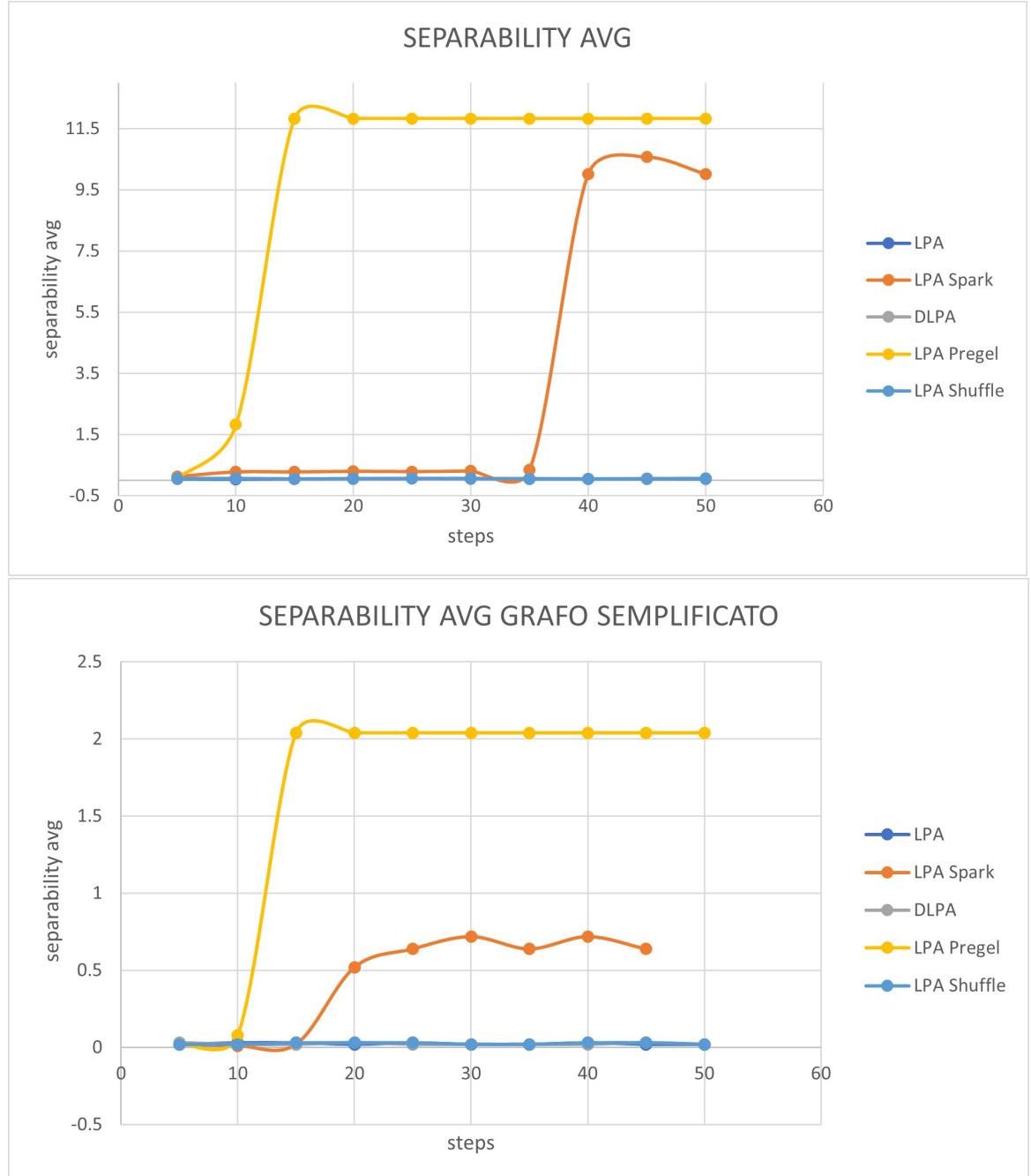
- DLPA dovrebbe essere un'implementazione specifica per la tipologia del nostro grafo, grafo diretto, e quindi restituire un numero minore di community e ognuna di esse dovrebbe essere di maggiore qualità. Questo però non ha evidenze nei nostri risultati essendo il numero di community individuate da DLPA molto simili a quello individuato da LPA.

8.3 Confronto modularità



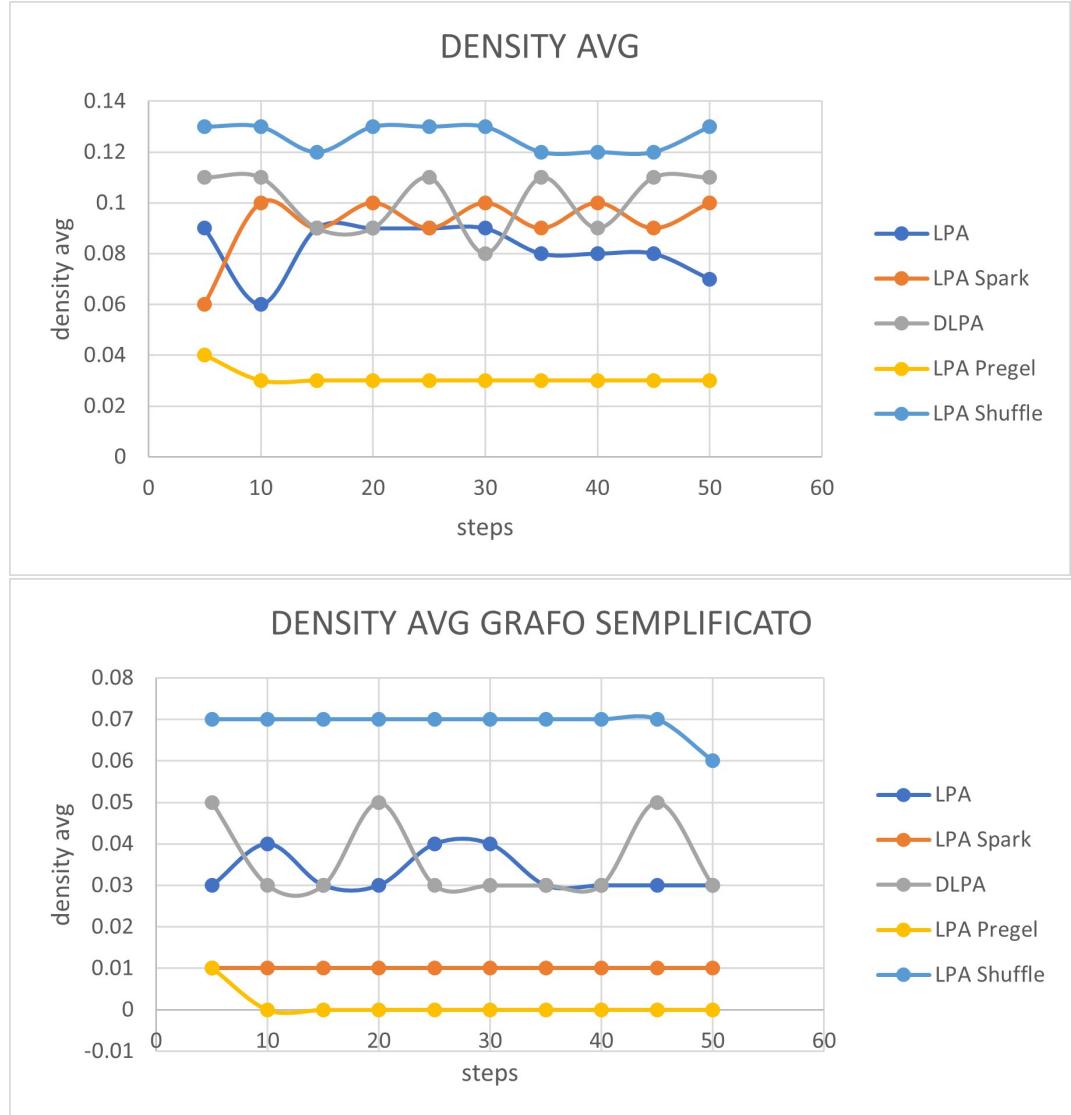
Dal punto di vista della modularità, l'algoritmo che sembra essere più stabile riguardo a tale metrica risulta essere: LPA Shuffle.

8.4 Confronto separability



Dal punto di vista della separability, l'algoritmo che sembra essere più stabile riguardo a tale metrica risulta essere: LPA Pregel.

8.5 Confronto density



Dal punto di vista della density, l'algoritmo che sembra essere più stabile riguardo a tale metrica risulta essere: LPA Shuffle.