

Meetup: analisi della partecipazione agli eventi

Davide Agnoletto Giovanni Toto

15 dicembre 2020



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Meetup è un servizio di rete sociale che ha lo scopo di facilitare l'incontro di gruppi di persone in varie località del mondo.

Meetup consente ai membri di trovare e unirsi a gruppi creati attorno a un comune interesse.



Meetup è un servizio di rete sociale che ha lo scopo di facilitare l'incontro di gruppi di persone in varie località del mondo.

Meetup consente ai membri di trovare e unirsi a gruppi creati attorno a un comune interesse.

Un utente può inserire la località del proprio domicilio/residenza e l'argomento di proprio interesse per visualizzare i gruppi legati a quella località e argomento.



Ogni utente è iscritto a uno o più gruppi.

Ogni gruppo può organizzare degli eventi a cui possono partecipare gli iscritti al gruppo.

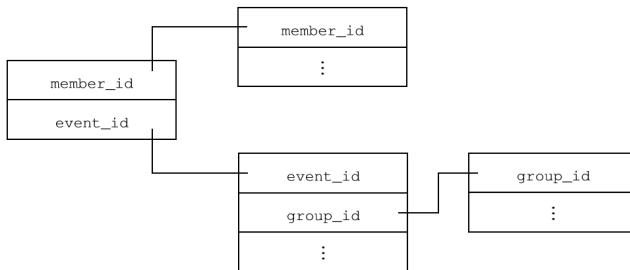
Ogni gruppo tratta un argomento, che fa riferimento ad una categoria più generale.

Il dataset *Nashville Meetup Network*, ottenuto tramite il sito *Kaggle*, è stato generato durante l'evento *Principles of Network Analysis with NetworkX* e utilizzato in una serie di tutorial presentati nei gruppi *PyNash* e *PyTennessee*.

Il dataset utilizzato contiene le partecipazioni agli eventi organizzati a Nashville dal 2015 al 2017 e fa riferimento a 24 476 utenti e 19 031 eventi.



Il dataset è formato da un serie di tabelle, una di riferimento e tre di metadati, collegate attraverso le variabili identificative di utenti, eventi, gruppi e categorie:



Per svolgere le analisi, si sono uniti i dataset attraverso le variabili identificative.

partecipazioni	
member_id	codice identificativo dell'utente
event_id	codice identificativo dell'evento
membri	
member_id	codice identificativo dell'utente
member_name	nome e cognome
hometown	città di nascita
city	città di residenza
state	stato della città di residenza
lat	latitudine dell'abitazione
lon	longitudine dell'abitazione
eventi	
event_id	codice identificativo dell'evento
event_name	nome dell'evento
time	data e ora dell'evento
group_id	codice identificativo del gruppo che l'ha organizzato
gruppi	
group_id	codice identificativo del gruppo
group_name	nome del gruppo
num_members	membri del gruppo
category_id	codice identificativo della categoria di appartenenza
category_name	nome della categoria di appartenenza
organizer_id	codice identificativo dell'organizzatore
group_urlname	url del gruppo nel sito

partecipazioni	
member_id	codice identificativo dell'utente
event_id	codice identificativo dell'evento
membri	
member_id	codice identificativo dell'utente
member_name	nome e cognome
hometown	città di nascita
city	città di residenza
state	stato della città di residenza
lat	latitudine dell'abitazione
lon	longitudine dell'abitazione
eventi	
event_id	codice identificativo dell'evento
event_name	nome dell'evento
time	data e ora dell'evento
group_id	codice identificativo del gruppo che l'ha organizzato
gruppi	
group_id	codice identificativo del gruppo
group_name	nome del gruppo
num_members	membri del gruppo
category_id	codice identificativo della categoria di appartenenza
category_name	nome della categoria di appartenenza
organizer_id	codice identificativo dell'organizzatore
group_urlname	url del gruppo nel sito

Si propongono due approcci per analizzare il fenomeno delle partecipazioni agli eventi:

- 1** Analisi dei legami tra eventi

Si propongono due approcci per analizzare il fenomeno delle partecipazioni agli eventi:

- 1 Analisi dei legami tra eventi
- 2 Analisi della partecipazione al singolo evento

Il legame tra una coppia di eventi è definito come il numero di utenti che hanno partecipato a entrambi gli eventi.

L'analisi dei partecipanti comuni permette di studiare l'interesse della popolazione nei confronti di determinati gruppi e/o categorie.

Il legame tra una coppia di eventi è definito come il numero di utenti che hanno partecipato a entrambi gli eventi.

L'analisi dei partecipanti comuni permette di studiare l'interesse della popolazione nei confronti di determinati gruppi e/o categorie.

In particolare, si è interessati a identificare quali gruppi e/o categorie sono più connesse tra loro e determinare la struttura delle relazioni che li lega.

Per l'analisi dei partecipanti comuni si utilizzano tre strumenti:

- Analisi descrittiva della rete di eventi
- Modelli per dati di rete
- Analisi delle associazioni tra categorie

Si considera la rete degli eventi in cui:

- Un nodo corrisponde ad un evento.
- Un arco tra due nodi esiste se i due eventi hanno almeno un partecipante in comune; l'arco è pesato con il numero di partecipanti in comune.

Nodi	19 031
Archi	1 438 370
Densità	0.0079
Categorie	31
Gruppi	602
Utenti distinti	24 631
Partecipazioni totali	126 813
Partecipazioni medie	5.15

La rete presenta un numero troppo elevato di nodi, oltre a una densità estremamente bassa.

Per motivi sia computazionali sia di visualizzazione e interpretazione, si decide di considerare un sottoinsieme di nodi.

Per fare ciò si ritiene adeguato considerare solo gli eventi con il maggior numero di partecipanti, in particolare si individua un *soglia* al di sotto della quale un evento viene ignorato.

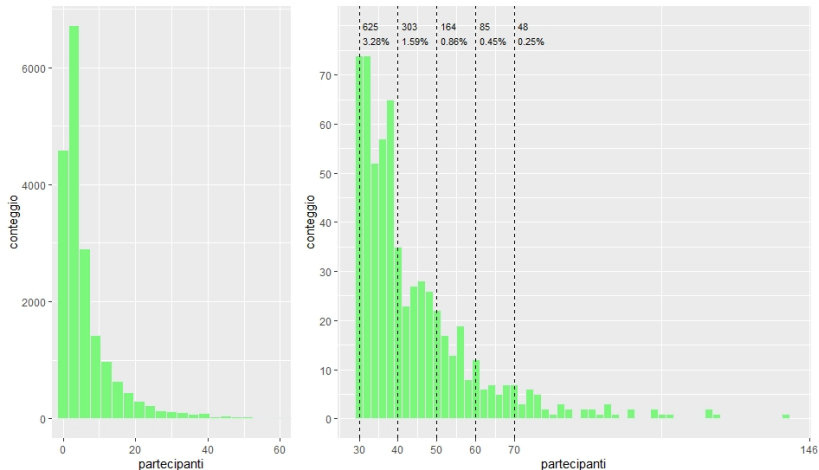


Figura: Frequenza di eventi per numero di partecipanti.

	soglia				
	70	60	50	40	30
Tech	42	63	108	179	324
Socializing	2	10	18	35	65
Career & Business	2	7	15	34	81
Outdoors & Adventures	1	3	6	11	15
Movies & Film	1	2	5	11	16
Religion & Beliefs	0	0	9	20	31
Dancing	0	0	2	7	25
Pet & Animals	0	0	1	2	12
Singles	0	0	0	2	7
Games	0	0	0	0	31
Other	0	0	0	2	18
Total	48	85	164	303	625

Tabella: Numero di eventi con almeno *soglia* partecipanti per categoria.

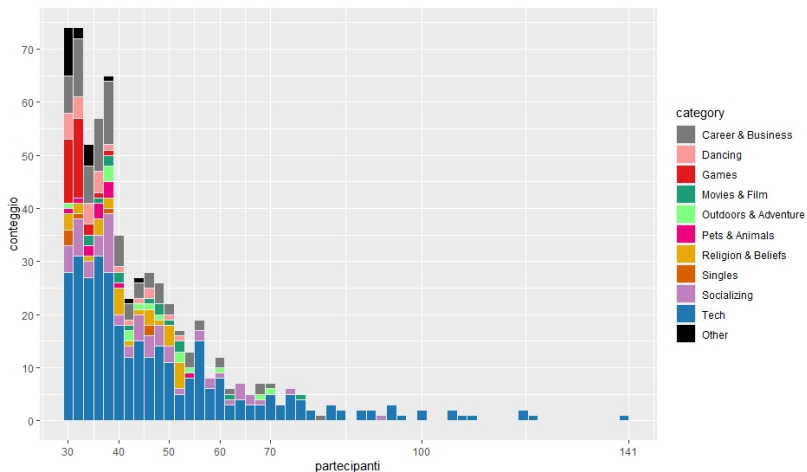
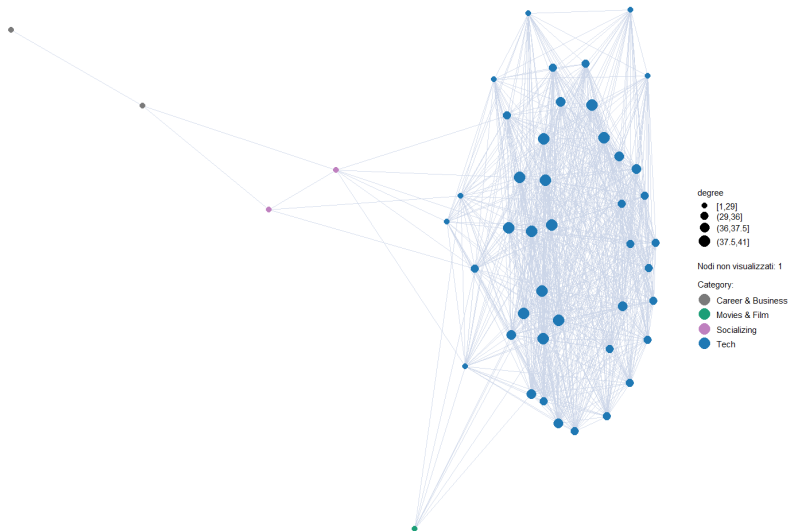
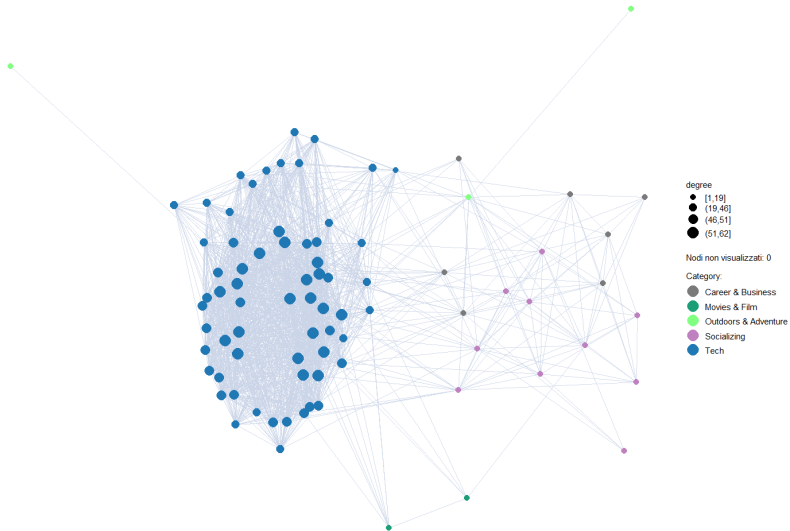


Figura: Frequenza di eventi per numero di partecipanti.

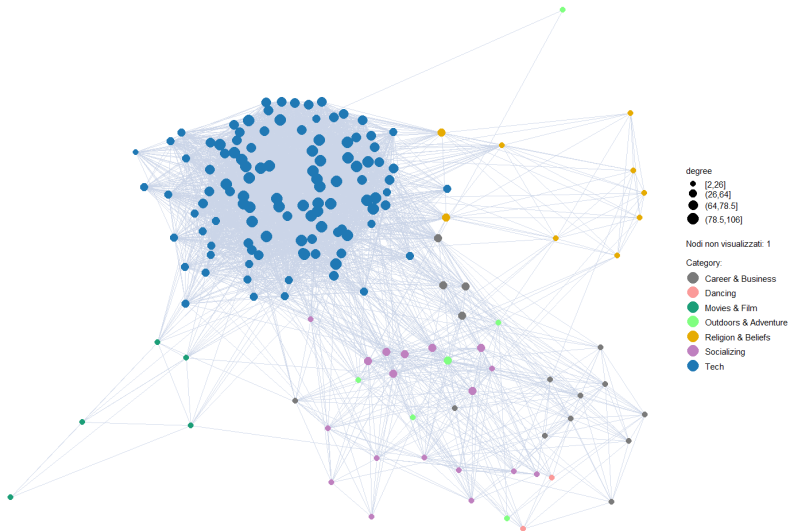
Grafo con soglia 70



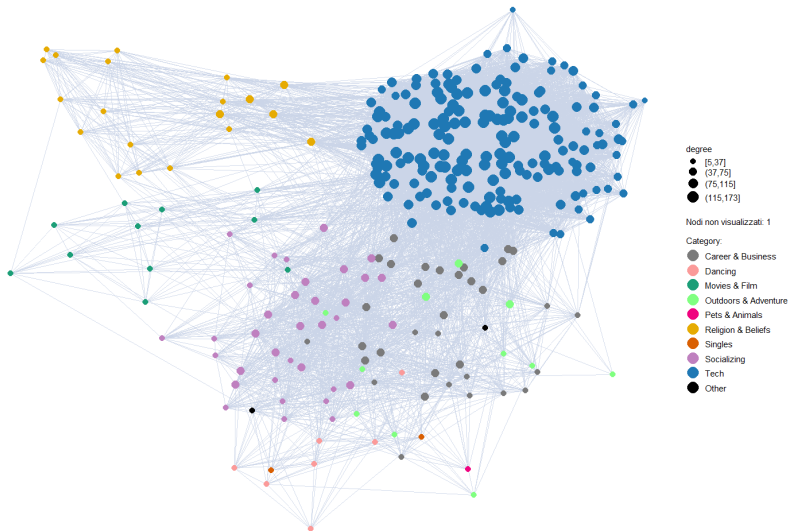
Grafo con soglia 60



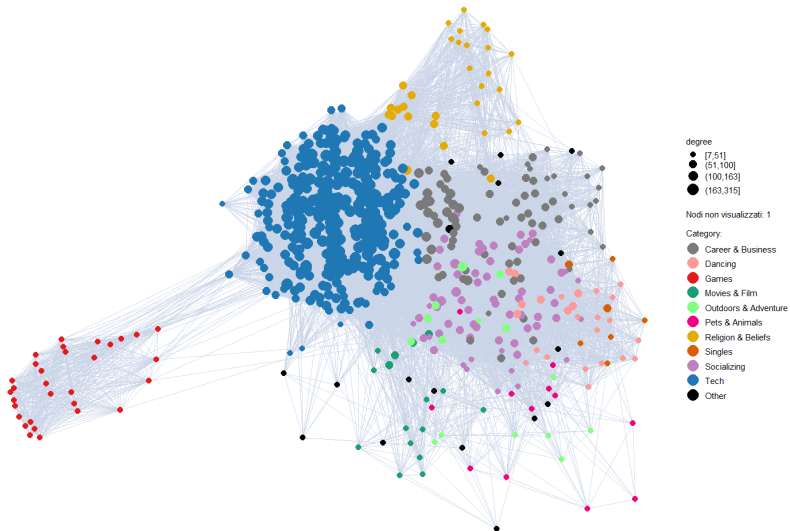
Grafo con soglia 50



Grafo con soglia 40



Grafo con soglia 30



	soglia				
	70	60	50	40	30
nodi	48	85	164	303	625
archi	740	1 602	4 428	11 761	34 577
densità	0.66	0.45	0.33	0.26	0.18
grado medio	30.83	37.69	54	77.63	110.65
betweenness media	23.68	45.58	86.87	161.73	361.57
diametro	20	8	24	21	28
categorie	5	5	8	10	18
gruppi	12	23	38	50	77
utenti distinti	2 219	3 423	5 151	6 945	9 703
partecipazioni totali	4 166	6 535	10 779	16 934	27 978
partecipazioni medie	1.88	1.91	2.09	2.44	2.88

Tabella: Statistiche descrittive per la rete al variare della *soglia*.

Si stima un *Additive and Multiplicative Effects Model*.

Si considerano solo gli eventi con almeno 50 partecipanti.

Le variabili di nodo, uguali per riga e colonna, sono:

- categoria di appartenenza
- numero di iscritti al gruppo di appartenenza

Le variabili di nodo, uguali per riga e colonna, sono:

- categoria di appartenenza
- numero di iscritti al gruppo di appartenenza

Le variabili diadiche sono:

- appartenenza allo stesso gruppo
- appartenenza alla stessa categoria

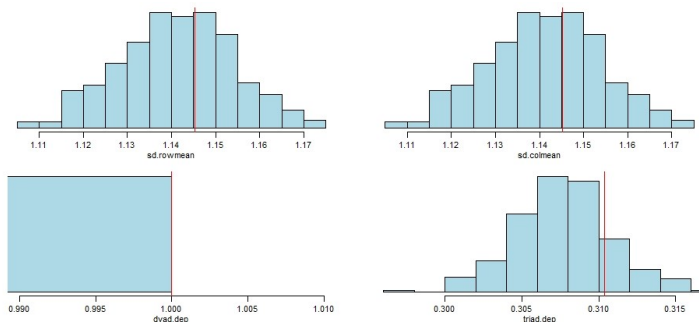


Figura: Risultati del miglior modello, ovvero un *modello AME* con un effetto moltiplicativo latente unidimensionale ($R=1$); si considerano solo gli eventi con almeno 50 partecipanti.

Regression coefficients:

	pmean	psd	z-stat	p-val
intercept	-0.052	0.204	-0.254	0.800
Movies_&_Film.node	0.223	0.181	1.231	0.218
Outdoors_&_Adventure.node	-0.268	0.314	-0.853	0.394
Pets_&_Animals.node	-0.005	0.372	-0.014	0.989
Religion_&_Beliefs.node	0.467	0.160	2.923	0.003
Socializing.node	-0.157	0.129	-1.216	0.224
Tech.node	0.124	0.104	1.202	0.230
Dancing.node	-0.001	0.264	-0.004	0.997
membri_iscritti.node	0.000	0.000	0.332	0.740
same_group.dyad	13.114	0.071	183.484	0.000
same_category.dyad	1.085	0.048	22.619	0.000

Variance parameters:

	pmean	psd
va	0.124	0.017
ve	2.047	0.019

Analisi delle associazioni tra categorie



lhs		rhs	supporto	fiducia	lift
{Career & Business,Outdoors & Adventure}	⇒	{Socializing}	0.0062	0.4872	3.2008
{Food & Drink}	⇒	{Socializing}	0.0117	0.3273	2.1503
{Singles}	⇒	{Socializing}	0.0083	0.3040	1.9974
{Singles}	⇒	{Outdoors & Adventure}	0.0070	0.2578	1.6490
{Food & Drink}	⇒	{Outdoors & Adventure}	0.0092	0.2559	1.6370
{Career & Business,Socializing}	⇒	{Outdoors & Adventure}	0.0062	0.2550	1.6312
{Dancing}	⇒	{Socializing}	0.0111	0.2549	1.6746
{Singles}	⇒	{Dancing}	0.0068	0.2489	5.7025
{Outdoors & Adventure,Socializing}	⇒	{Career & Business}	0.0062	0.2460	1.3946
{Dancing}	⇒	{Outdoors & Adventure}	0.0104	0.2372	1.5172
{Health & Wellbeing}	⇒	{New Age & Spirituality}	0.0098	0.2269	4.8227
{Music}	⇒	{Socializing}	0.0093	0.2178	1.4307
{New Age & Spirituality}	⇒	{Health & Wellbeing}	0.0098	0.2079	4.8227
{Community & Environment}	⇒	{Outdoors & Adventure}	0.0056	0.2032	1.2998
{Language & Ethnic Identity}	⇒	{Career & Business}	0.0061	0.1871	1.0610

Osservando i grafici al variare della *soglia* e le associazioni identificate, si ritiene interessante svolgere ulteriori analisi su alcuni cluster:

- 1 *Tech*
- 2 *Socialità* = {*Career & Business*
Outdoors & Adventure
Socializing}

Osservando i grafici al variare della *soglia* e le associazioni identificate, si ritiene interessante svolgere ulteriori analisi su alcuni cluster:

- 1 *Tech*
- 2 *Socialità* = {*Career & Business*
Outdoors & Adventure
Socializing}

Mantenendo la struttura della rete e l'idea di filtraggio dei nodi in base ad una soglia, si ripetono le stesse analisi nei due cluster identificati.

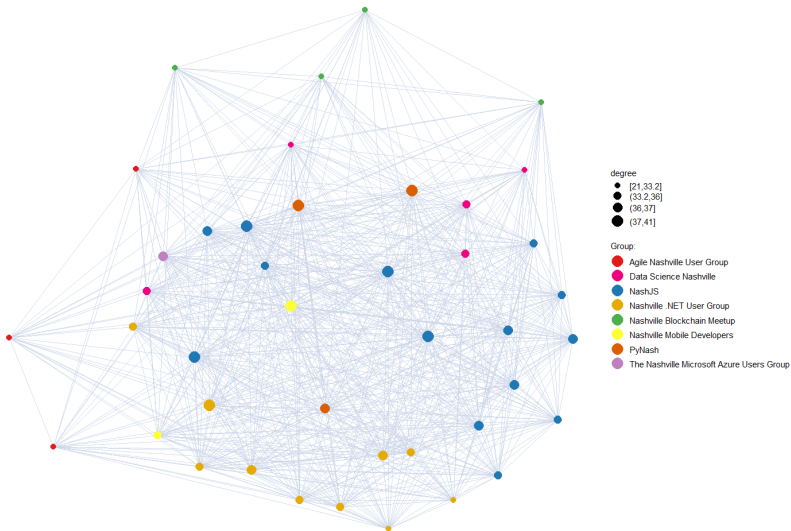
Cluster Tech: gruppi rilevanti



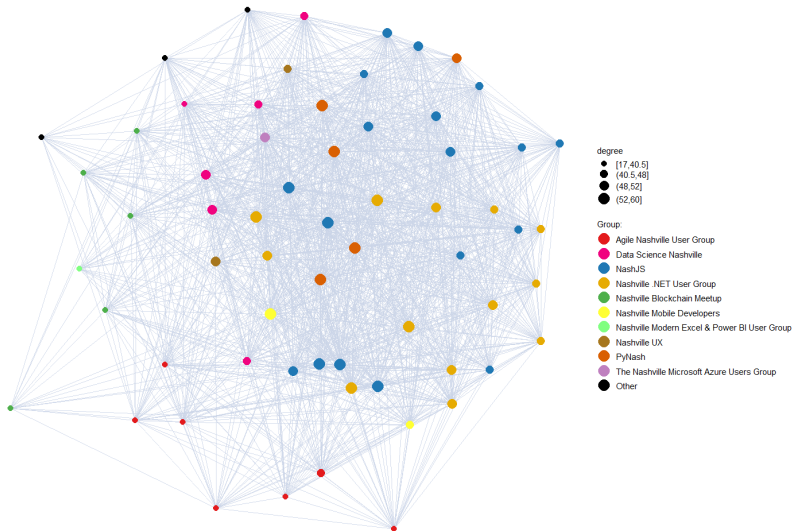
	soglia					
	70	60	50	40	30	20
NashJS	14	18	21	21	23	23
Nashville .NET User Group	10	13	17	17	21	22
Data Science Nashville	5	6	7	11	13	15
Nashville Blockchain Meetup	4	5	6	9	10	10
Agile Nashville User Group	3	7	11	15	20	22
PyNash	3	5	9	16	19	24
Nashville Mobile Developers	2	2	2	5	8	13
The Nashville Microsoft Azure Users Group	1	1	1	1	6	11
Nashville UX	0	2	10	16	21	29
Nashville Modern Excel & Power BI User Group	0	1	4	12	17	22
NashBI	0	1	2	2	7	8
Greater Nashville Healthcare Analytics	0	1	1	1	3	9
Nashville Bloggers	0	1	1	1	1	2
Nashville DevOps Meetup	0	0	3	11	16	18
Code for Nashville	0	0	3	5	12	29
Nashville Product Meetup	0	0	2	6	15	23
Nashville Machine Learning Meetup	0	0	2	3	5	6
Nashville API Developers	0	0	2	3	4	4
Franklin Developer Lunch & Learn	0	0	1	11	18	20
Nashville Java Users' Group	0	0	1	3	8	13
NashReact	0	0	1	2	8	14
Design Thinking Nashville	0	0	1	2	5	11
Total	42	63	108	179	324	539

Tabella: Numero di eventi con almeno *soglia* partecipanti per gruppo.

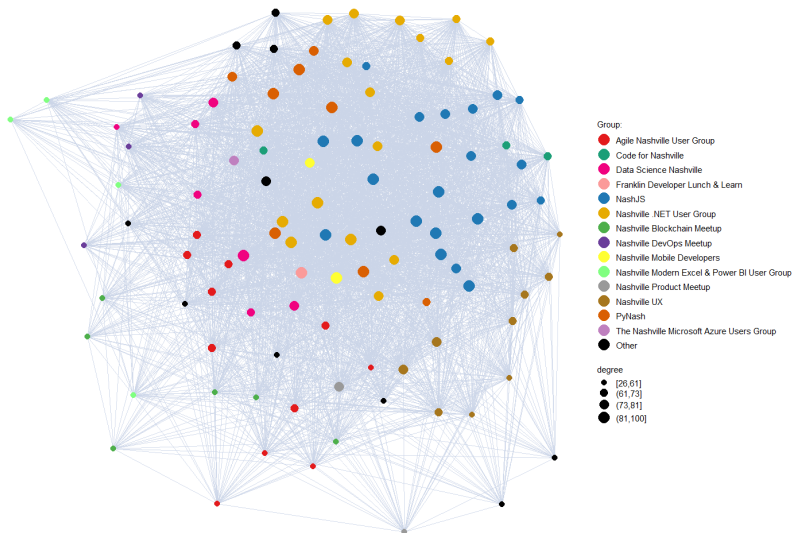
Cluster *Tech*: grafo con soglia 70



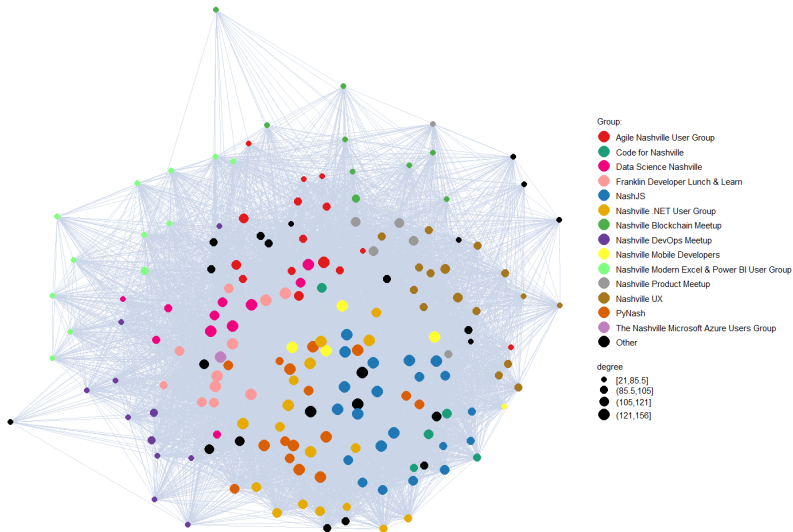
Cluster Tech: grafo con soglia 60



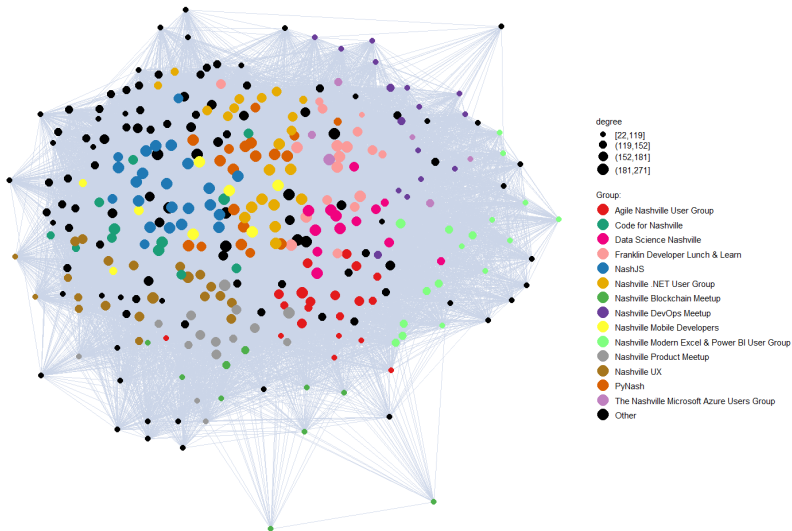
Cluster Tech: grafo con soglia 50



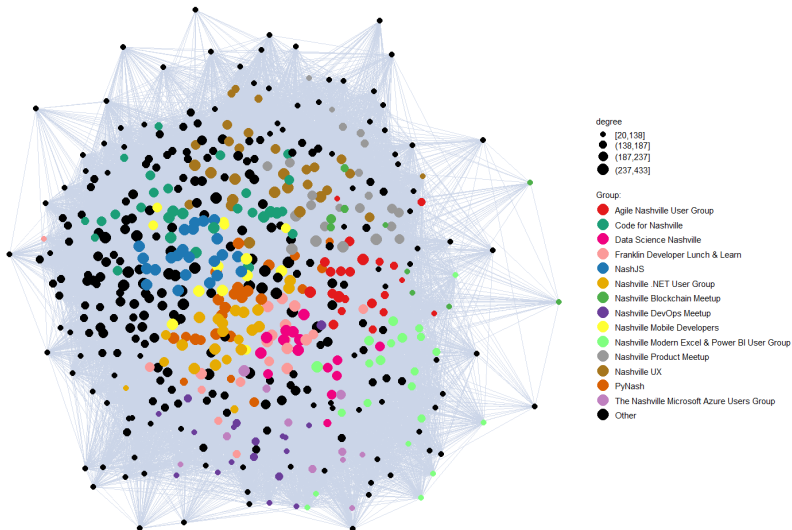
Cluster Tech: grafo con soglia 40



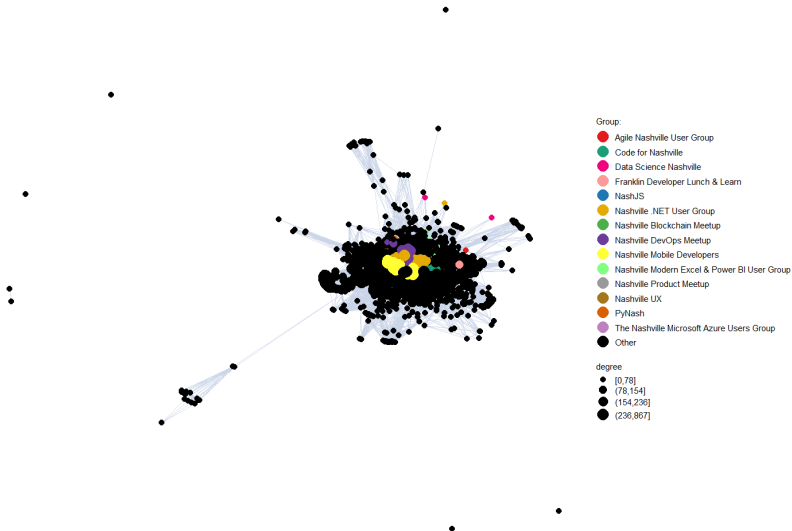
Cluster Tech: grafo con soglia 30



Cluster Tech: grafo con soglia 20



Cluster *Tech*: grafo senza soglia



	soglia					
	70	60	50	40	30	20
nodi	42	63	108	179	324	539
archi	720	1444	3727	9070	24267	51322
densità	0.84	0.74	0.64	0.57	0.46	0.35
grado medio	34.29	45.84	69.02	101.34	149.8	190.43
betweenness media	17.88	24.84	40.27	66.27	122.63	212.9
diametro	4	3	3	3	3	3
gruppi	8	13	22	28	42	57
utenti distinti	1797	2243	2996	3600	4585	5367
partecipazioni totali	3700	5036	7483	10622	15630	20793
partecipazioni medie	2.06	2.25	2.5	2.95	3.41	3.87

Tabella: Statistiche descrittive al variare della *soglia*.

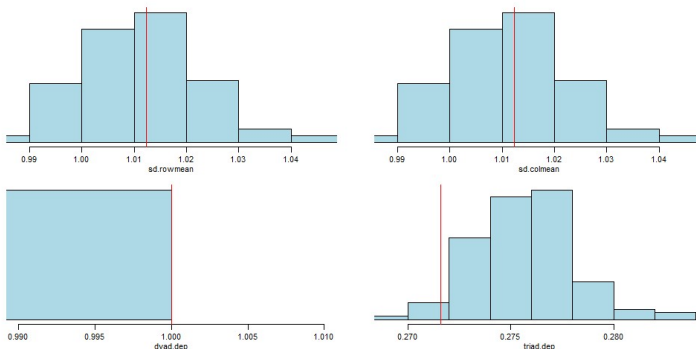


Figura: Risultati del miglior modello, ovvero un *modello AME* con un effetto moltiplicativo latente bidimensionale ($R=2$); si considerano solo gli eventi con almeno 40 partecipanti.

Regression coefficients:

	pmean	psd	z-stat	p-val
intercept	1.094	0.330	3.316	0.001
DataScienceNashville.node	0.356	0.161	2.219	0.026
FranklinDeveloper.node	0.162	0.162	0.998	0.318
NashJS.node	0.627	0.202	3.098	0.002
.NetUserGroup.node	0.330	0.163	2.018	0.044
DevOpsMeetup.node	-0.156	0.158	-0.984	0.325
ModernExcelPowerBIUser.node	-0.321	0.160	-2.002	0.045
NashvilleUX.node	-0.431	0.152	-2.833	0.005
PyNash.node	0.285	0.160	1.777	0.076
altro_gruppo.node	0.071	0.117	0.609	0.542
membri_iscritti.node	0.000	0.000	1.027	0.304
same_group.dyad	12.021	0.049	245.556	0.000

Variance parameters:

	pmean	psd
va	0.143	0.018
ve	1.969	0.017

lhs		rhs	supporto	fiducia	lift
{Code for Nashville}	⇒	{NashJS}	0.0866	0.4442	1.4244
{Code for Nashville,PyNash}	⇒	{NashJS}	0.0283	0.5897	1.8911
{NashJS,PyNash}	⇒	{Code for Nashville}	0.0283	0.4340	2.2265
{Code for Nashville,Nashville .NET User Group}	⇒	{NashJS}	0.0267	0.6915	2.2173
{NashJS,Nashville .NET User Group}	⇒	{Code for Nashville}	0.0267	0.4276	2.1940
{Code for Nashville,Nashville UX}	⇒	{NashJS}	0.0263	0.7619	2.4431
{NashJS,Nashville UX}	⇒	{Code for Nashville}	0.0263	0.4384	2.2490
{Code for Nashville,The Iron Yard - Nashville}	⇒	{NashJS}	0.0226	0.7237	2.3206
{NashJS,The Iron Yard - Nashville}	⇒	{Code for Nashville}	0.0226	0.4661	2.3913
{Nashville UX,The Iron Yard - Nashville}	⇒	{NashJS}	0.0172	0.6364	2.0406
{Nashville UX,PyNash}	⇒	{NashJS}	0.0160	0.7647	2.4521
{Nashville .NET User Group,PyNash}	⇒	{NashJS}	0.0156	0.5938	1.9039
{Nashville .NET User Group,Nashville UX}	⇒	{NashJS}	0.0148	0.6667	2.1377
{Nashville UX,The Iron Yard - Nashville}	⇒	{Code for Nashville}	0.0127	0.4697	2.4098
{Code for Nashville,The Iron Yard - Nashville}	⇒	{Nashville UX}	0.0127	0.4079	1.8970

Cluster *Socialità*: gruppi rilevanti



	soglia					
	70	60	50	40	30	20
Transplant Nashville	2	5	8	11	14	16
WOMEN "Word of Mouth Entrepreneurial Networkers"	2	4	7	12	18	22
Nashville Flight Training	1	1	1	1	1	1
20's & 30's Women looking for girlfriends	0	3	6	11	16	27
Nashville Hiking Meetup	0	2	5	7	9	24
Nashville SEO & Internet Marketing, Over 1,600 Members!	0	2	2	6	14	20
20s in Nashville	0	2	3	7	20	48
Music City Young Professionals	0	1	1	4	6	6
Nashville Young Professionals Meetup	0	1	2	5	13	16
Nashville Online Entrepreneurs	0	0	2	6	13	17
Business Girls Rock! Nashville Entrepreneurship Meetup	0	0	1	1	2	6
Women 'n' Wine of Williamson County	0	0	1	1	9	18
Paddle Adventures Unlimited	0	0	0	3	4	6
Eat Love Nash	0	0	0	2	9	25
Nashville Professional Referral Club TEAM Chapter	0	0	0	2	3	5
Nashville Network After Work - Business Networking Events	0	0	0	1	2	5
Nashville's Society of Women Business Owners (SOWBO)	0	0	0	0	4	11
Women's Business Network	0	0	0	0	1	9
Nashville Backpacker Meetup	0	0	0	0	1	11
Nashville Online Business Strategy Meetup	0	0	0	0	1	6
Total	5	20	39	80	161	353

Tabella: Numero di eventi con almeno *soglia* partecipanti per gruppo.

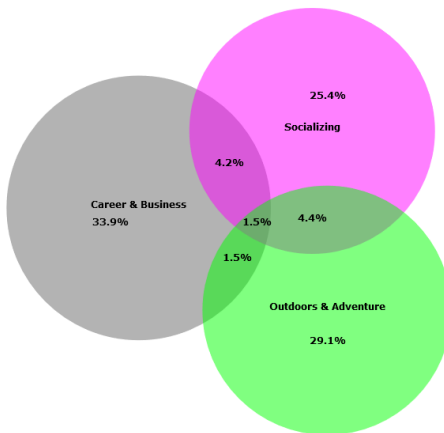
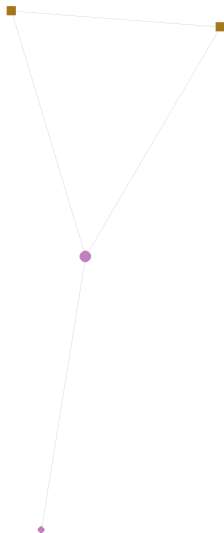


Figura: Rappresentazione dei partecipanti in comune fra le tre categoria.

Cluster *Socialità*: grafo con soglia 70



Category:

- Career & Business
- ▲ Outdoors & Adventure
- Socializing

degree

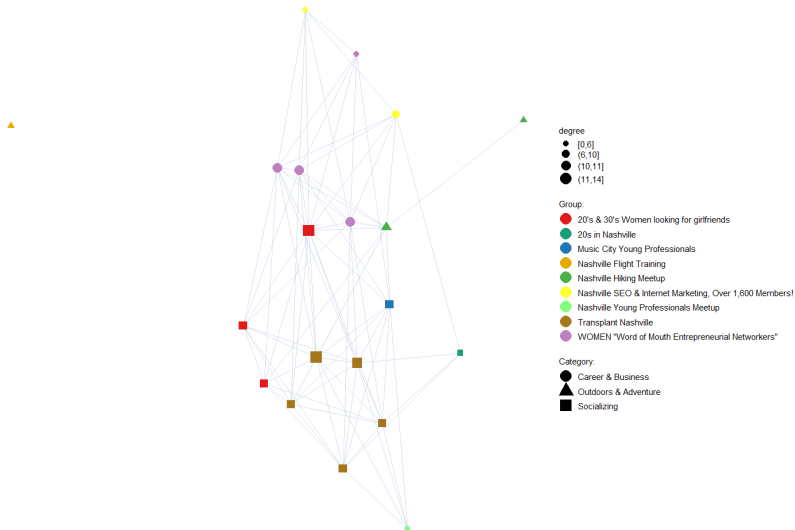
- [0, 1]
- (1, 2]
- (2, 3]

Group:

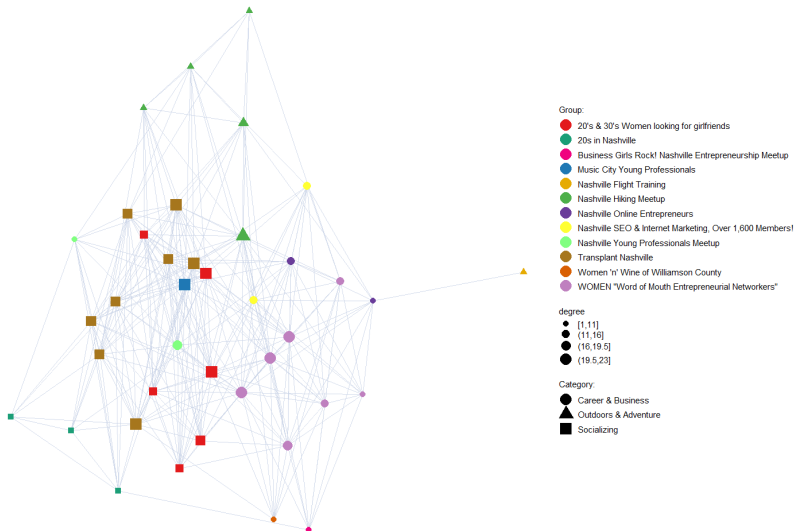
- Nashville Flight Training
- Transplant Nashville
- WOMEN "Word of Mouth Entrepreneurial Networkers"



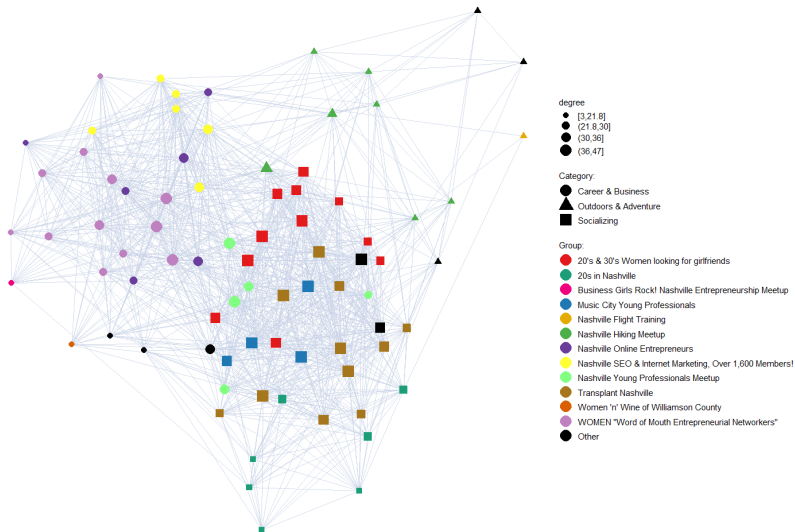
Cluster *Socialità*: grafo con soglia 60



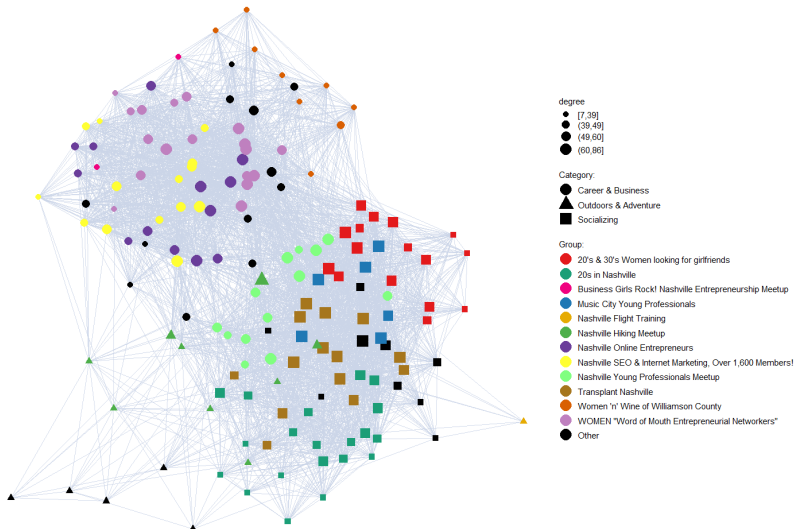
Cluster *Socialità*: grafo con soglia 50



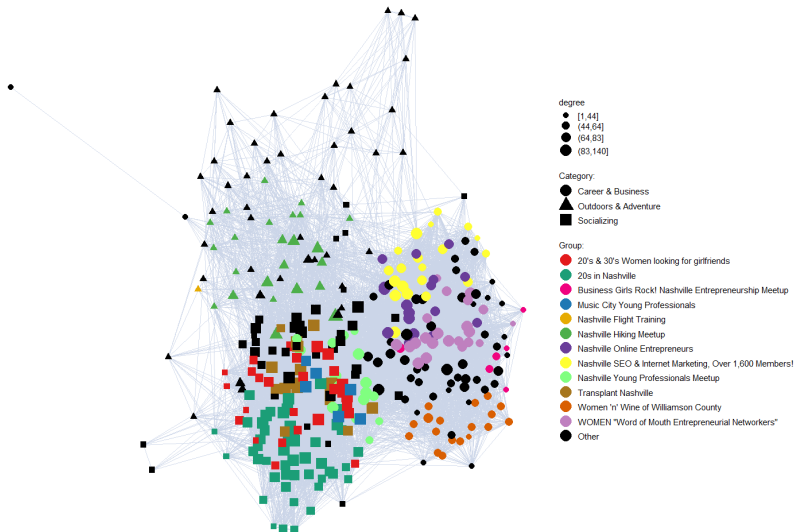
Cluster *Socialità*: grafo con soglia 40



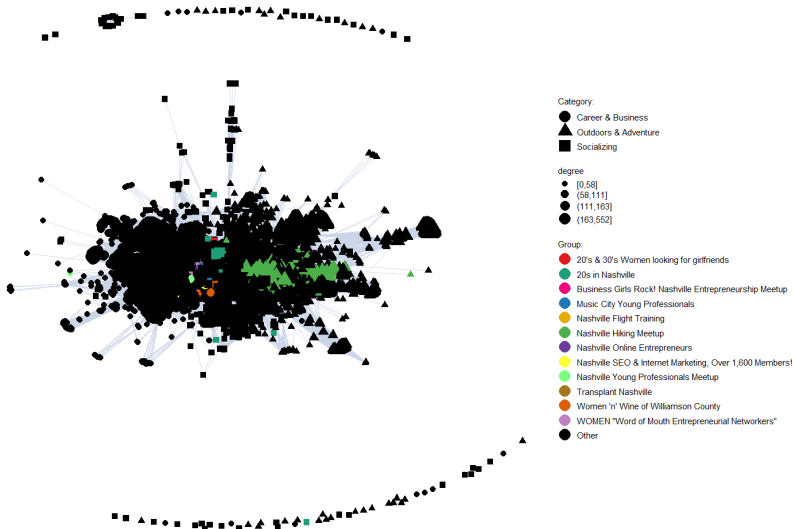
Cluster *Socialità*: grafo con soglia 30



Cluster *Socialità*: grafo con soglia 20



Cluster *Socialità*: grafo senza soglia



	soglia					
	70	60	50	40	30	20
nodi	5	20	39	80	161	353
archi	4	83	293	1153	3913	11073
densità	0.4	0.44	0.4	0.36	0.3	0.18
grado medio	1.6	8.3	15.03	28.82	48.61	62.74
betweenness media	0.6	8.12	19.29	36.87	74.1	186.56
diametro	17	8	4	3	3	13
gruppi	3	9	12	16	20	40
utenti distinti	358	1093	1748	2605	3659	5322
partecipazioni totali	390	1360	2379	4210	7056	11570
partecipazioni medie	1.09	1.24	1.36	1.62	1.93	2.17

Tabella: Statistiche descrittive al variare della *soglia*.

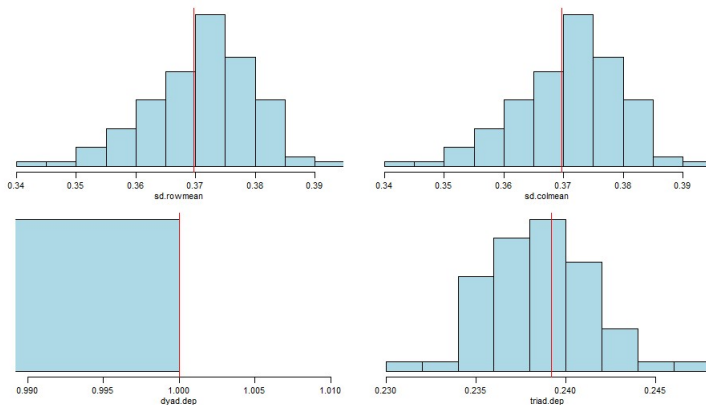


Figura: Risultati del miglior modello, ovvero un *modello AME* con un effetto moltiplicativo latente bidimensionale ($R=2$); si considerano solo gli eventi con almeno 30 partecipanti.

Regression coefficients:

	pmean	psd	z-stat	p-val
intercept	0.072	0.100	7.352	0.000
Outdoors_&_Adventure.node	-0.205	0.156	-1.316	0.188
Socializing.node	-0.082	0.073	-1.125	0.260
membri_iscritti.node	0.000	0.000	-1.844	0.065
same_group.dyad	5.120	0.036	142.599	0.000
same_category.dyad	-0.016	0.037	-0.440	0.660

Variance parameters:

	pmean	psd
va	0.093	0.012
ve	0.747	0.007



L'analisi precedente permetteva di valutare esclusivamente l'affluenza comune a due eventi, ma non si focalizzava sulla partecipazione al singolo evento.

L'analisi precedente permetteva di valutare esclusivamente l'affluenza comune a due eventi, ma non si focalizzava sulla partecipazione al singolo evento.

Si è ora interessati a prevedere la partecipazione al singolo evento e, se possibile, identificare le caratteristiche che la influenzano.

L'analisi precedente permetteva di valutare esclusivamente l'affluenza comune a due eventi, ma non si focalizzava sulla partecipazione al singolo evento.

Si è ora interessati a prevedere la partecipazione al singolo evento e, se possibile, identificare le caratteristiche che la influenzano.

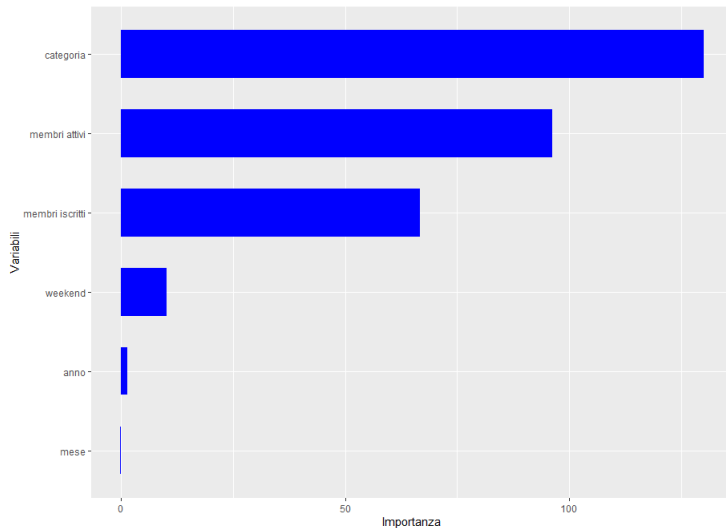
Si applicano i modelli dell'inferenza classica a tutti gli eventi del dataset e, successivamente, solo agli eventi appartenenti ai due cluster individuati in precedenza.

Le variabili considerate sono:

- categoria di appartenenza
- numero di iscritti al gruppo di appartenenza
- membri considerati attivi (almeno una partecipazione)
- anno in cui si è tenuto l'evento
- mese in cui si è tenuto l'evento
- weekend (indica se l'evento si tiene durante il weekend o meno)

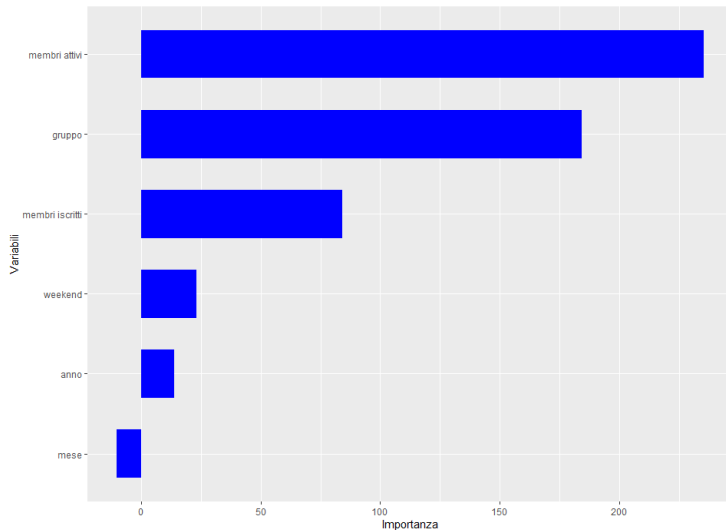
Modello	MAE
modello lineare	4.50
regressione lasso	4.47
mars	4.54
albero di regressione	4.22
rete neurale	4.23
gradient boosting	3.42
random forest	3.15

Output random forest



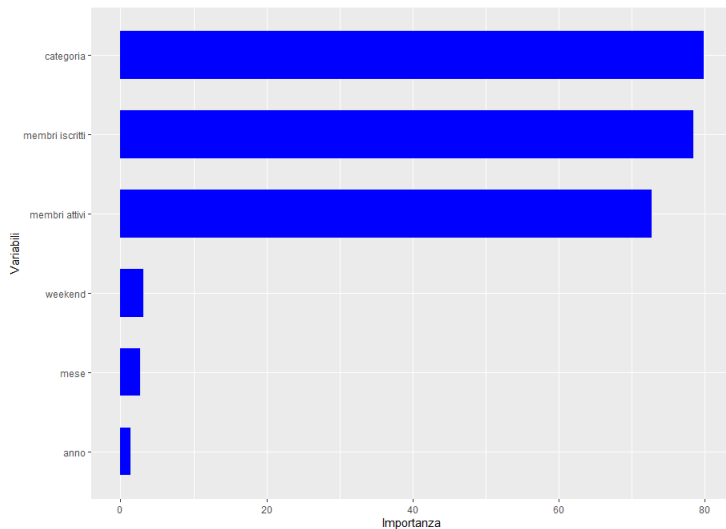
Modello	MAE
modello lineare	7.60
regressione lasso	7.67
mars	8.42
albero di regressione	7.38
rete neurale	7.91
gradient boosting	7.22
random forest	6.81

Output random forest - *Tech*



Modello	MAE
modello lineare	4.62
regressione lasso	4.62
mars	4.45
albero di regressione	4.20
rete neurale	4.85
gradient boosting	3.83
random forest	3.60

Output random forest - *Socialità*



Riepilogando, sono state effettuate due analisi parallele che hanno approcciato il problema in due modi diversi:

Riepilogando, sono state effettuate due analisi parallele che hanno approcciato il problema in due modi diversi:

1 Analisi dei legami tra eventi

Si è interessati a identificare quali gruppi e/o categorie sono più connesse tra loro e determinare la struttura delle relazioni che li lega.

Riepilogando, sono state effettuate due analisi parallele che hanno approcciato il problema in due modi diversi:

1 Analisi dei legami tra eventi

Si è interessati a identificare quali gruppi e/o categorie sono più connesse tra loro e determinare la struttura delle relazioni che li lega.

2 Analisi della partecipazione al singolo evento

Si è interessati a prevedere la partecipazione al singolo evento e, se possibile, ad identificare le caratteristiche che la influenzano.

Fra gli eventi a maggior partecipazione, si sono identificati quattro cluster, corrispondenti a diverse categorie:

- *{Tech}*
- *{Games}*
- *{Religion & Beliefs}*
- *{Career & Business, Outdoors & Adventure, Socializing, ...}*

Fra gli eventi a maggior partecipazione, si sono identificati quattro cluster, corrispondenti a diverse categorie:

- *{Tech}*
- *{Games}*
- *{Religion & Beliefs}*
- *{Career & Business, Outdoors & Adventure, Socializing, ...}*

Gli eventi tendono ad avere più partecipanti in comune con altri eventi dello stesso cluster, a prescindere dal gruppo di appartenenza, e averne meno con eventi di altri cluster come si evinceva anche nelle analisi associative.

{Tech}

- Cluster con eventi a più alta partecipazione e in genere molti partecipanti in comune.
- Come ci si aspetta, eventi dello stesso gruppo tendono ad avere più partecipanti in comune ma ciò risulta meno evidente se si considerano solo eventi ad alta affluenza.
- I gruppi più centrali sono *NashJS*, *PyNash* e *Nashville .NET User Group*; in particolare, il primo raccoglie utenti di diversi gruppi con argomenti spesso molto diversi tra loro.

{*Career & Business, Outdoors & Adventure, Socializing, ...*}

- Sebbene il cluster contenga 28 categorie, ci si è concentrati solo su quelle sopra riportate poiché sono quelle con più eventi a più alta partecipazione.
- La maggior parte degli eventi hanno meno di 40 partecipanti.
- L'appartenenza allo stesso gruppo ha un minore effetto sulla partecipazione comune rispetto al cluster *Tech*.
- Se si considerano solo eventi ad alta affluenza la distinzione fra le tre categorie di eventi sembra meno netta rispetto a quando si considerano anche eventi con meno partecipanti.

Il modello migliore risulta essere sempre la random forest che non è interpretabile ma fornisce una misura di importanza delle variabili.

In questa situazione, si ritiene che la precisione previsiva sia più importante dell'interpretabilità poiché un'interpretazione del fenomeno è già fornita dal primo approccio.

Il secondo approccio può essere quindi considerato complementare al primo.

Ci si è concentrati su determinati cluster poiché si è preferito dare maggior importanza a quelli con un maggior numero di eventi ad alta affluenza: ovviamente è possibile effettuare un'analisi analoga ai cluster rimanenti.

Data la struttura dei dati, è possibile impostare altre analisi focalizzandosi sugli utenti e non più sugli eventi.