

Progetto ICT

Il network che andremo ad analizzare è ricavato da un dataset contenente un campione di imprese marchegiane operanti in diversi settori, suddetto dataset è stato costruito dal professor Alessandro Marra. Il network tratta di affinità / prossimità tecnologica tra le imprese. Le imprese, che costituiscono i nodi, presentano un attributo dato dal settore ateco a due cifre, tuttavia, anche se sono presenti codici ateco differenti, e quindi diversi settori di mercato, il network risultante comprenderà tutte le aziende prese in esame. Le imprese che avranno un link tra loro sono quelle che condividono almeno 3 tecnologie sulle 10 considerate (e.g., Big Data, Internet of Things, robot etc).

Per prima cosa importiamo su Gephi i due file excel contenenti i nodi (Node table network of technological proximity) e gli archi (Edge table network of technological proximity) controllando che Gephi stia interpretando correttamente il contenuto e facendo attenzione a spuntare l'opzione "append to existing workspace" per il secondo dataset.

Il grafico ottenuto (FIG. 1) non sembra essere in grado di fornirci informazioni.

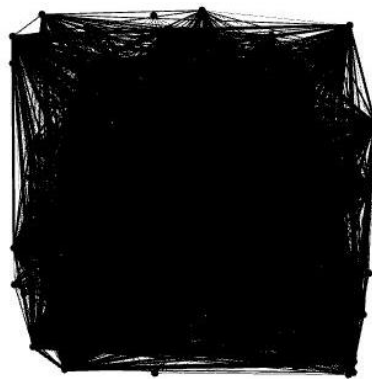


FIG. 1

Procediamo quindi all'applicazione di un layout. In particolare, andremo ad utilizzare un force-based layout, nello specifico Yifan Hu. Come possibile vedere dalla FIG. 2 il grafico, dopo avervi applicato il layout, risulta ancora poco comprensibile ma sicuramente meglio distribuito.

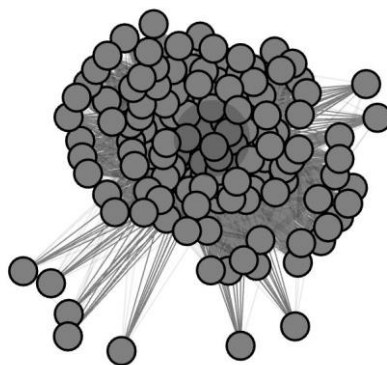


FIG. 2

Continuiamo cercando di rendere il nostro grafico più fruibile, cominciamo riducendo lo spessore degli archi, distinguendo i nodi associando loro colori differenti in base ai rispettivi codici ateco, infine ridimensionando la grandezza dei nodi sulla base della loro *degree centrality* (calcolata tramite la funzione *average degree* all'interno della sezione statistics di Gephi). Come sappiamo questa statistica ci permette di assegnare ad ogni nodo un valore relativo al numero di nodi ad esso collegati, più alto sarà il numero di

archi collegati ad un nodo più alto sarà il valore a questo assegnato. La media che vediamo è relativamente alta, parliamo infatti di una degree centrality media pari a 85,7. Il grafico diventa adesso più comprensibile come è possibile constatare nella FIG. 3. Osservando il network notiamo che molti nodi sono concentrati al centro del grafo questo probabilmente è legato all'elevato valore dell'average degree.

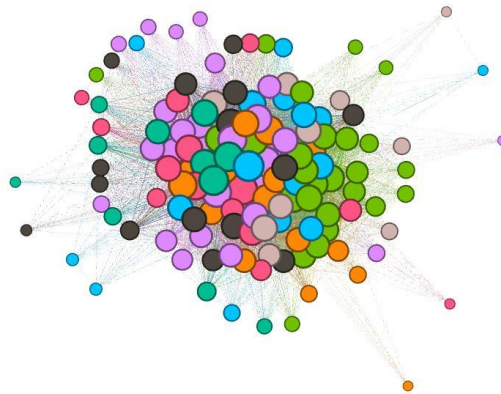


FIG. 3

Possiamo cominciare a ipotizzare che nonostante i codici ateco siano differenti le tecnologie utilizzate dalle imprese marchigiane siano simili tra le varie aziende, infatti non notiamo una separazione troppo evidente tra i vari settori. Il network sembra ricco di connessioni, sono ben poche le osservazioni isolate.

Abbiamo poi deciso di colorare i nodi in base alla modularity class, ma, poiché questa si attesta ad un valore pari a 0.083 che testimonia una bassa propensione del network ad essere diviso in *cluster* abbiamo ritenuto che una visualizzazione basata su questa statistica non sia da prendere in considerazione poiché i cluster risultanti non sarebbero particolarmente significativi. Gephi in ogni caso ci propone una divisione in quattro cluster, come possiamo vedere in FIG. 4. Vediamo che c'è una forte distinzione tra due classi (blu e arancione), mentre le altre due classi non sono ben distinte.-----

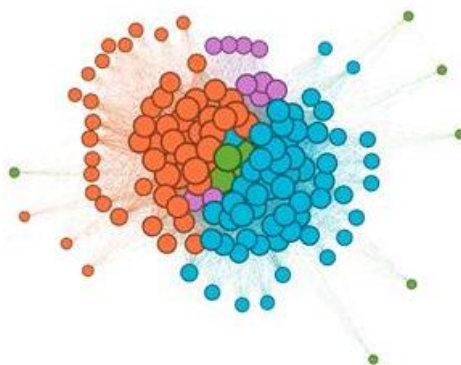


FIG. 4

Calcolo delle principali network statistics

Centrality measures

L'average degree nel nostro caso è come avevamo detto di 85,72, questo significa che un'impresa condivide almeno tre tecnologie in media con altre 86 imprese. Questo valore appare abbastanza alto considerando che il numero totale di nodi è 137. Dividendo l'average degree per il numero di nodi otteniamo il valore standardizzato di 0,62 il quale ci indica che un'impresa utilizza le stesse tecnologie di circa il 62% delle imprese presenti sul territorio. Osservando il grafico della degree distribution non siamo in

grado di riconoscere alcun pattern, né *bell-shaped* né *long-tailed*. Calcolando poi la *Average Weighted Degree* arriviamo ad avere un valore di 337,97 ben quattro volte il valore calcolato da Gephi per l'*average degree*, questo ci porta ad ipotizzare che nel nostro grafo siano presenti alcuni nodi con un numero elevato di archi a loro connessi. Guardando i "Weight" nell'*edge table* vediamo che ci sono sei archi di peso massimo, ovvero quelli risultanti da una condivisione di dieci tecnologie su dieci.

Nella sezione data laboratory possiamo osservare una grande variabilità dei degree collegati ai singoli nodi, si passa da 10 a 136 quindi alcune imprese saranno poco collegate alle altre dal punto di vista tecnologico mentre altre, come quelle verdi centrali che abbiamo visto nella modularity class sono collegate a tutte le altre. Sono presenti quattro nodi con degree centrality massima ovvero 136, due di questi sono appartenenti alle classi ateco 25 mentre gli altri due alle classi ateco 28 e 46.

Questi quattro sono anche i nodi con "Closeness Centrality" più alta (pari a 1) proprio perché sono connessi a tutti gli altri nodi; la Closeness Centrality minima è invece pari a 0.52.

Questi nodi sono anche quelli con betweenness centrality più alta (151), questo mostra che hanno un ruolo di gatekeeper (o bridge) e che un alto numero di shortest path passa attraverso loro. La betweenness centrality più bassa è pari a 0 e appartiene tendenzialmente a nodi lontani dal centro del grafico.

L' *Eccentricity* raggiunge il valore massimo di 2, questo è sempre dettato dalla presenza dei quattro nodi che sono connessi a tutti. Tutti i nodi diversi dai quattro con degree centrality più alta avranno quindi eccentricity pari a 2 poiché passeranno per questi per raggiungere gli altri nodi.

Network diameter

Questo parametro rileva il numero massimo di step necessari per attraversare il network da parte a parte, il valore assunto in questo caso è pari a 2, particolarmente basso se consideriamo la *6- degree separation theory*.

Graph Density

Il 63% circa dei nodi sono tra loro connessi, sembra essere un numero elevato il che ci può far pensare che questo network sia di tipo "SMALL WORLD"

Average (shortest) path length

L'average (shortest) path length è 1,37, in media quindi per giungere da un nodo all'altro è necessario percorrere meno di un due archi.

Connected Components

Questo parametro è pari a 1, un valore quantomeno atteso poiché abbiamo nodi con eccentricity pari a 1.

Average clustering coefficient

Nel nostro network è pari a 0,868, indica che più di 4/5 dei triangoli sono chiusi. Il nostro *local clustering coefficient* varia da 1 a 0,62 quindi, nel caso peggiore, ci ritroviamo ad avere nodi con il 62% di triangoli, di cui quel nodo è vertice, chiusi.