CENTRALITA' nella RETE

Nozione di Centralità

- Nozione introdotta dai sociologi nel dopoguerra (Bavelas 1948)
- L'ipotesi è che la centralità strutturale sia un elemento in grado di motivare l'importanza di un attore in un processo di comunicazione
- Ma tra i sociologi non esisteva consenso su questa ipotesi e mancava una definizione precisa della nozione di centralità
- Freeman ne propone una rifacendosi alla teoria dei grafi

Centrality in Social Networks: conceptual clarification, Linton Freeman, 1979.

Centralità nelle reti

- Nello studio delle reti complesse la nozione di centralità può essere importante per:
  - giudicare la rilevanza/criticità di nodi o aree delle rete
  - attribuire una misura di distanza fra nodi o aree delle rete
  - identificare il grado di coesione di un area delle rete
  - identificare le aree di una rete (i gruppi coesi, le sue comunità)

Centralità di Freeman

- Freeman identifica tre nozioni di centralità:
  - Una legata al grado di un nodo: **Degree Centrality**

$$C_D(n_i) = \sum_{k=1}^{N} a(n_i, n_k)(N-1)^{-1}$$

- Per un grafo G = (N, A) dove l'arco a_{ik} è contato pari a *1* quando sia presente un collegamento tra il nodo n_i e un nodo n_k

Centralità di Freeman

- Freeman identifica tre nozioni di centralità:
  - Una legata alle distanze tra i nodi: **Closeness Centrality**

$$C_C(n_i)^{-1} = \sum_{k=1}^{N} d(n_i, n_k)$$

- Dove *d* è il cammino minimo (o distanza geodesica) tra i nodi n_j e il nodo n_k

Centralità di Freeman

- Freeman identifica tre nozioni di centralità:
  - Una legata ai percorsi che collegano i nodi: **Betweenness Centrality**

$$C_B(n_i) = \sum_{j}^{N} \sum_{k}^{N-1} \frac{D_{jk}(n_i)}{D_{jk}}$$

- Dove D_{jk} è l'insieme di tutti i cammini minimi che collegano il nodo n_j e il nodo n_k
- Dove D_{jk} (n_i) è l'insieme di tutti i cammini minimi che collegano il nodo n_j e il nodo n_k passando per n_i

Comunità

- Le tre misure di centralità identificano anche tre definizioni diverse di aree della rete
- L'idea di base che un area sia una zona nella quale i nodi abbiano una coesione tra loro
  - coesione in base al numero di connessioni dirette
  - coesione in base al numero di connessioni a una certa distanza
  - coesione in base alla resistenza alla soppressione di nodi

Componente

- Si può parlare di una **componente** di un grafo G = (N, A) se esiste un sottografo i cui nodi S $\subset$ N e gli archi A_S connettono i nodi S e non connettono i nodi S con nessun nodo non in S

Densità

- Si può parlare di **densità** di un grafo G = (N, A) come di misura del rapporto tra gli archi in A e il massimo numero di archi possibili tra i nodi di N

- Reti non direzionate (grafo non orientato)

$$\frac{2|A|}{N(N-1)}$$

- Reti direzionate (grafo orientato)

$$\frac{|A|}{N(N-1)}$$

- La densità è pari a 1 se tutti i nodi sono connessi tra loro, a volte chiamata **clique**

Clique

- La densità è pari a 1 se tutti i nodi sono connessi tra loro, a volte chiamata **clique**
- Si può rilassare questa nozione definendo una distanza massima *d* alla quale i nodi devo essere connessi
  - Una **d-clique** è il sottografo di massima estensione nel quale la distanza più ampia tra due nodi è pari a *d*
- Si può rilassare questa nozione definendo un numero *k* di nodi adiacenti minimo
  - Una **k-clique** (k-plex) è il sottografo di massima estensione nel quel ogni nodo ha un numero minimo di nodi adiacenti par a *k*

Comunità - Cluster …

- Si può parlare di una **cluster** di un grafo G = (N, A) se esiste un sottografo i cui nodi S $\subset$ N e gli archi A_S connettono i nodi S:
  - con una densità maggiore che i nodi non in S
    - quindi i numero di connessioni verso non S è minore che verso S
- Se S è un cluster allora tutti i sottografi del cluster hanno più connessioni coi loro complementi all'interno di S che con nodi all'esterno di S
  - si può quindi parlare di gerarchi di cluster

Algoritmi

- Chiaramente gli algoritmi per identificare i **cluster** non saranno una realizzazione delle definizione di base. Altrimenti la complessità sarebbe esponenziale
- Se un cluster è un area coesa della rete significa che quest'area resterà connessa anche a seguito della rimozione di alcuni nodi
- I nodi più importanti per la connessione dell'intera rete sono quelli con **betweenness centrality** più elevata

Algoritmi

- L'algoritmo base per l'identificazione di una gerarchia di cluster è dato da:

1) Calcolo di C_B per ogni *n*
2) Rimozione degli *n* con i valori più alti
3) Calcolo delle componenti
4) Ripetizione dei passi 1, 2 e 3 fino ad ottenere N componenti

    - Le componenti ottenute equivalgono ai cluster nella rete, la sequenza con cui sono state ottenute equivale alla loro gerarchia

Complessità

- Di base l'algoritmo ha una complessità pari a:

1) Calcolo di D_{jk} per ogni *n:* $O(AN)$

2) Calcolo di C_B per ogni *n:* $O(AN^2)$

3) Ripetizione dei passi fino ad ottenere N componenti: $O(AN^3)$

Similarità Strutturale

- Un altro modo per determinare se due nodi appartengono ad uno stesso gruppo è quello di valutarne la similarità strutturale

- Una tipica misura di similarità è la misura di Jaccard

$$S(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Similarità Strutturale

- Nel caso di una rete si può usare ad esempio per vedere quanti vicini (nodi adiacenti) sono condivisi da due nodi
  - Ad esempio definendo V(n) come l'insieme dei vicini di n:

$$S(A,B) = \frac{|V(n) \cap V(m)|}{|V(n) \cup V(m)|}$$

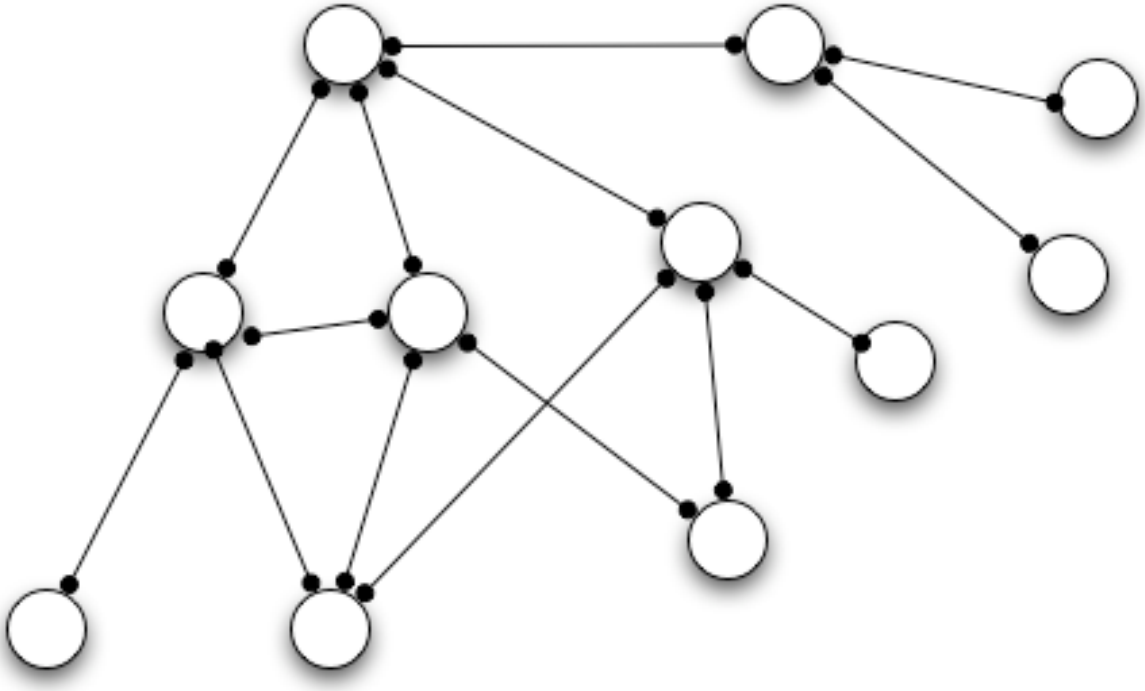
Similarità Strutturale

- Generalizzando invece di utilizzare una funzione particolare come quella per identificare i vicini di un nodo possiamo utilizzare una funzione generica

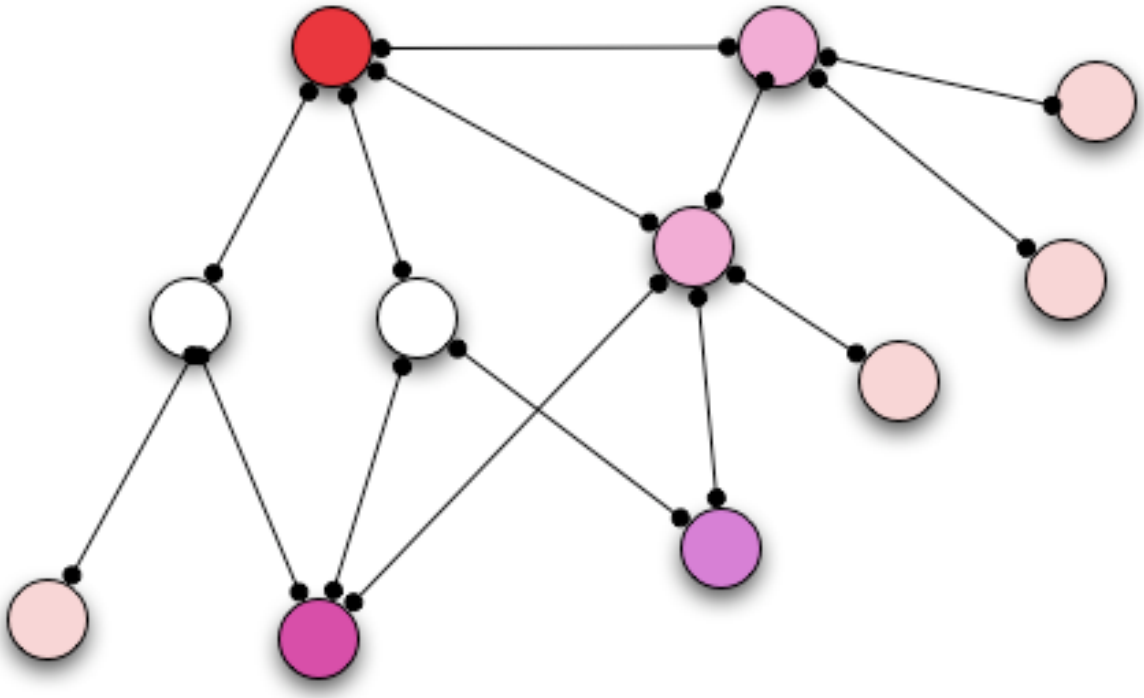
- Dovremo poi fornire un metodo per descrivere il comportamento di quella funzione

$$S(A,B) = \frac{|F(n) \cap F(m)|}{|F(n) \cup F(m)|}$$

Similarità Strutturale



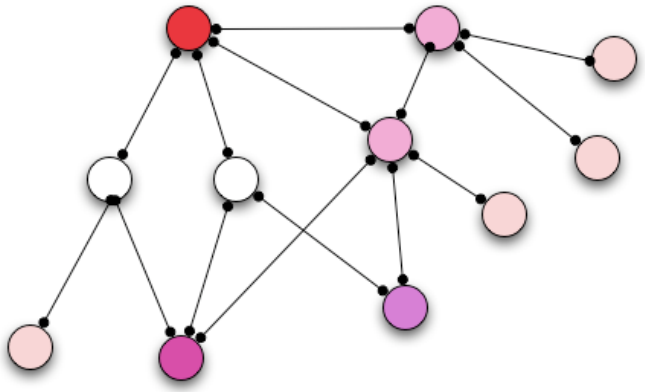
Similarità Strutturale



Similarità Strutturale

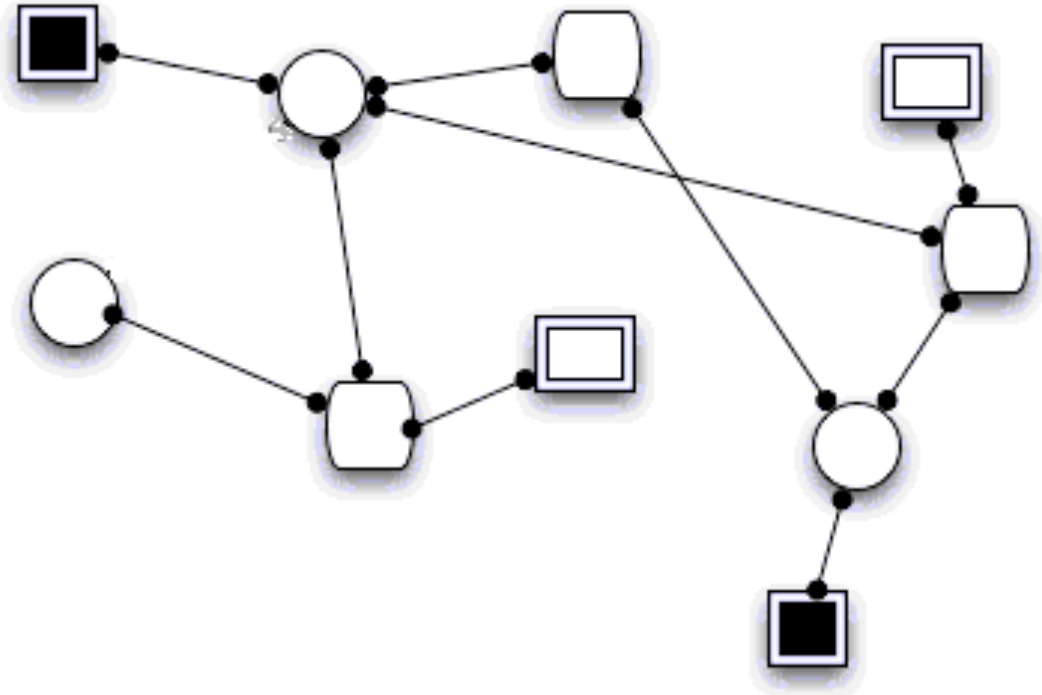
- In questo caso possiamo descrivere F(n) come l'insieme dei nodi a distanza 1

- La misura di similarità esprimerà quindi il rapporto tra il numero dei nodi adiacenti condivisi e tutti i distinti nodi adiacenti dei due nodi

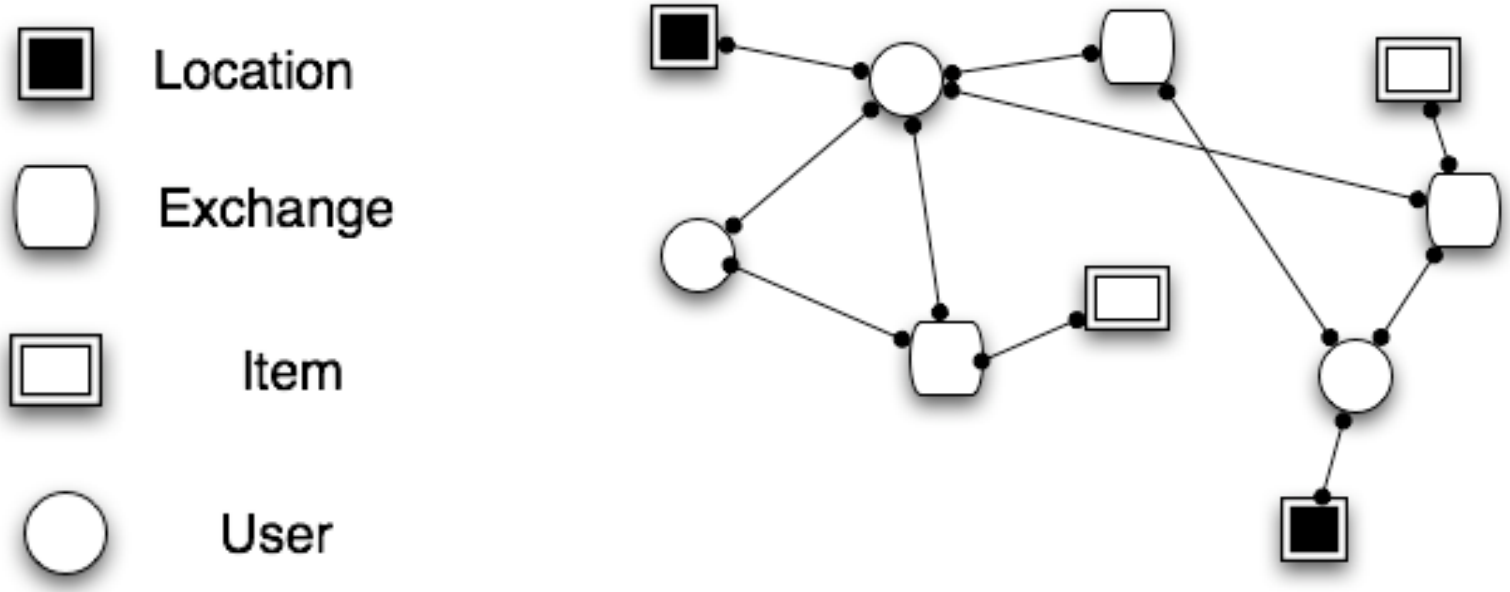


$$F(n) : \bigcup d(n,k) = 1$$

Similarità Strutturale

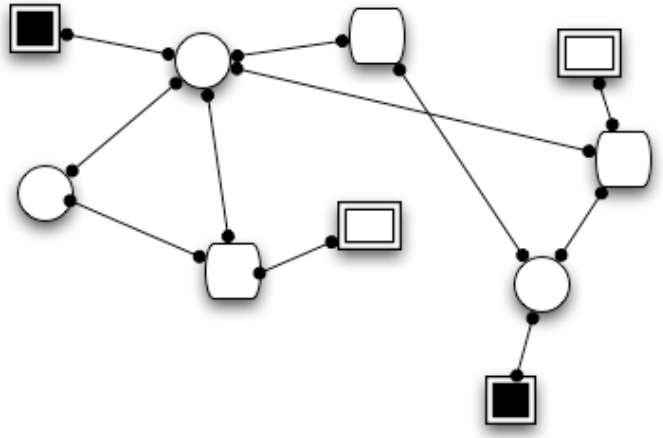


Similarità Strutturale



Similarità Strutturale

- In questo caso possiamo descrivere F(n) come una serie di vincoli espressi tramite condizioni sugli archi (predicati binari)



- Ad esempio potremmo dire che un utente scambia con utenti nella sua regione e che scambiano software

$$F(n) : l(n, R) \wedge i(n, S)$$

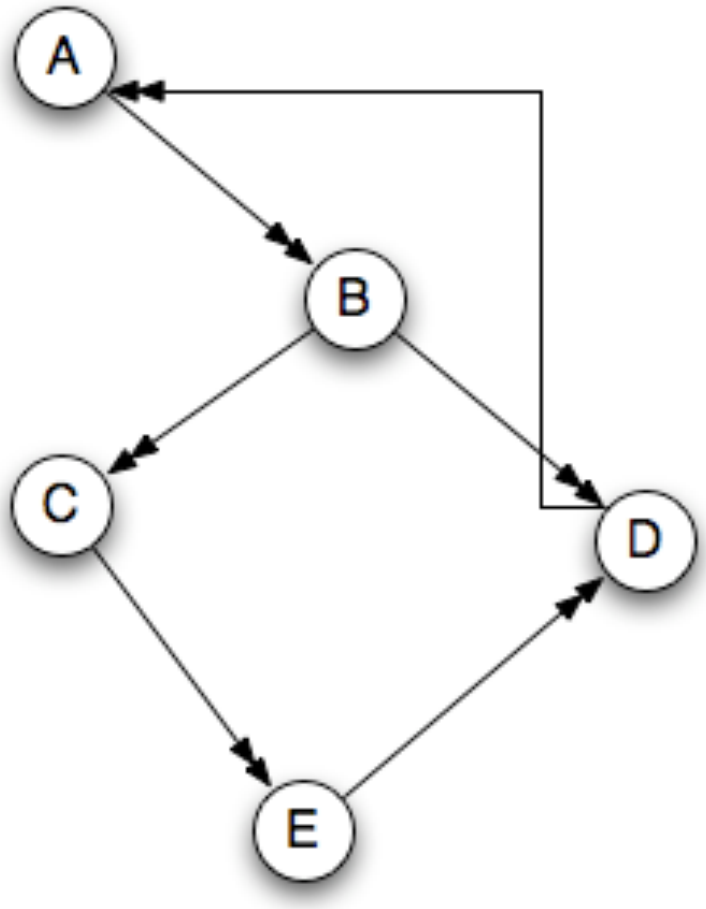
COSTRUZIONI

Matrici

- Dat grafo G = (N, A) è possibile descriverlo come una matrice di dimensione N x N dove verrà inserito un valore pari a per ogni posizione (j, k) quando sia presente un collegamento tra il nodo n_i e un nodo n_k

Matrici

	A	B	C	D	E
A		1			
B			1	1	
C					1
D	1				
E				1	



Matrici

- Attraverso la matrice si possono contare
  - in-degree: colonne
  - out-degree: righe
  - distanze tra due nodi
  - numero di cammini minimi
  - diametro
  - numero di loop

Matrici

- Per calcolare
  - distanze tra due nodi
  - numero di cammini minimi
  - diametro
  - numero di loop

è comodo definire una matrice delle distanze: per ogni nodo si listano gli archi a distanza da 1 fino al diametro, in pratica fino ad incontrare un loop oppure non poter più proseguire

Matrici

	d1	d2	d3	d4	d5
A	AB	BD BC	DA* CE	ED′	DA*
B	BD BC	DA CE	AB* ED′	DA′	AB*
C	CE	ED	DA	AB	BC* BD′
D	DA	AB	BD* BC	CE	ED*
E	ED	DA	AB	BD BC	DA′ CE*

Matrici

- Per calcolare
  - distanze tra due nodi
    - si parte da un nodo di inizio e si cerca in quale colonna della distanza è il nodo di destinazione
  - numero di cammini minimi
    - si leggono tutte le righe

Matrici

- Per calcolare
  - diametro
    - è la distanza massima
  - numero di loop
    - ad ogni colonan della distanza sono presenti i loop di quella lunghezza, per ottenere il numero di loop si divide per la lunghezza del loop

Clusterizzazione

- Usando la formula della **closness centrality** calcoliamo Cc per ogni nodo

$$A(1+2+2+3)^{-1} = \frac{1}{8}$$

$$B(1+1+2+2)^{-1} = \frac{1}{6}$$

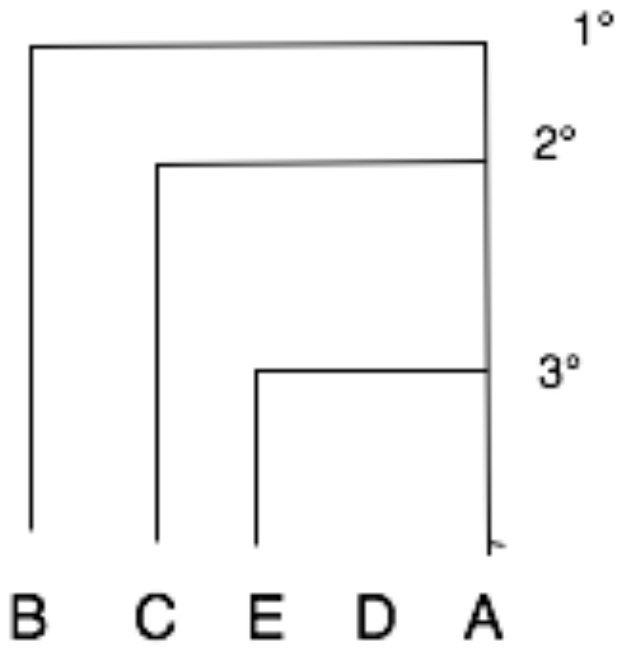
$$C(1+2+3+4)^{-1} = \frac{1}{10}$$

$$D(1+2+3+4)^{-1} = \frac{1}{10}$$

$$E(1+2+3+4)^{-1} = \frac{1}{10}$$

Clusterizzazione

- Possiamo costruire il dendogramma ordinando i cluster in base alle loro Cc
- Questi potranno essere definiti in base ad una misura di densità dei cammini minimi esistenti rispetto ai possibili (N*N-1)



Clusterizzazione

- Possiamo costruire il dendogramma ordinando i cluster in base alle loro Cc
- Questi potranno essere definiti in base ad una misura di densità dei cammini minimi esistenti rispetto ai possibili (N*N-1)