



POLITECNICO DI MILANO
Facoltà di Ingegneria dell'Informazione
Corso di Laurea Magistrale in Ingegneria Informatica

Sistema di Raccomandazione di Contatti Basato su Analisi Topologiche di una Rete Sociale

Relatore: Prof. Marco Colombetti

Correlatori: Ing. David Laniado
Ing. Riccardo Tasso

Tesi di Laurea di: Michele Monti - matricola 710251

Anno Accademico 2008/2009
Sessione di Laurea del 3 Maggio 2010

OBIETTIVO DELLA RICERCA

Mettere a punto delle tecniche che permettano di costruire un Sistema di Raccomandazione di contatti su una Rete Sociale che sia:

- Automatico
- Modulare
- Rapido

SISTEMI DI RACCOMANDAZIONE

- SCOPO: determinare un Ranking, utilizzando precisi algoritmi, che assegni importanza diversa a una serie di oggetti proposti all'utente.
- TIPOLOGIE:
 - Orientati ai contenuti
 - Collaborativi
 - Ibridi

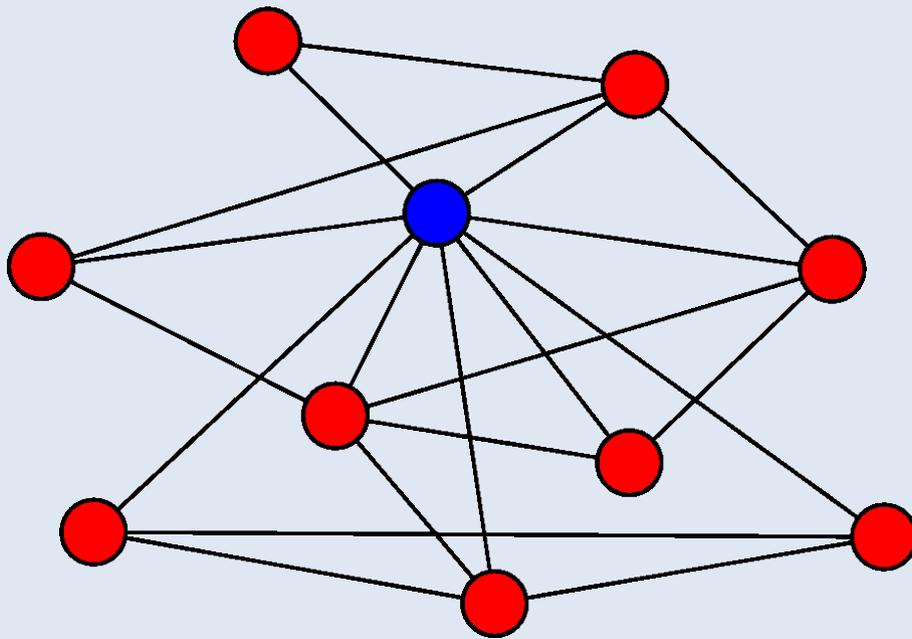
SISTEMI DI RACCOMANDAZIONE IN UNA RETE SOCIALE

- **CONTENT MATCHING:** controllo sui contenuti condivisi da differenti utenti, i quali potrebbero essere per questo messi in relazione di amicizia.
- **FRIENDS OF FRIENDS:** sfrutta la relazione transitiva tra terne di contatti.

*“se B è amico di A e C è amico di B
ALLORA può accadere che
 C sia amico di A ”*

PERSONAL NETWORK

È quella parte di una Rete Sociale che si incentra su un singolo utente (**NODO** blu) e su cui si definiscono tutte e sole le possibili Relazioni di Amicizia (**ARCHI** tra i due **NODI** rossi oppure **ARCHI** tra i **NODI** rossi e il blu).



È su di essa che vengono effettuate le **ANALISI TOPOLOGICHE**

CASO DI STUDIO

La Rete Sociale Online presa in esame è stata quella di FACEBOOK, sito gratuito e attivo dal 2004 e accessibile a chiunque.

CARATTERISTICHE:

- Il Profilo possiede le informazioni dell'utente e una bacheca su cui condividere contenuti, quali stati, foto, video, eventi e link.
- La **RELAZIONE di AMICIZIA** è SIMMETRICA.

METRICHE SEMPLICI

Cos'è un AMICO IN COMUNE?

Nodo che è all'interno della Personal Network e è punto di unione tra la PERSONA ESAMINATA e il CANDIDATO da SUGGERIRE.

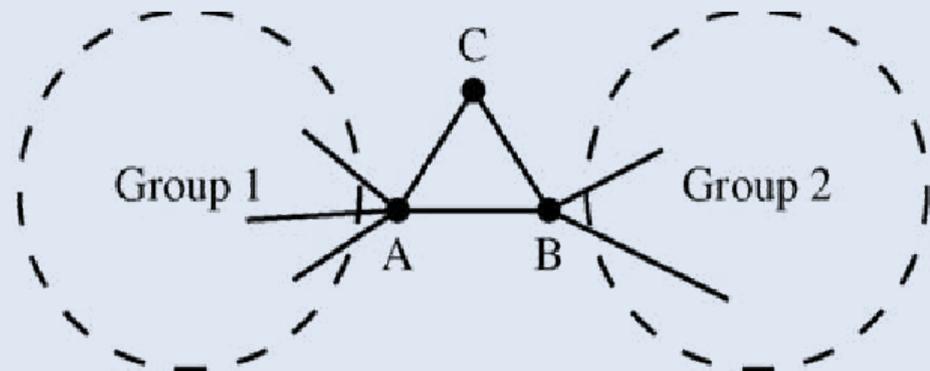
- *AMICI IN COMUNE (AC)*: suggerisce i candidati che hanno più AMICI in COMUNE.
- *AMICI IN COMUNE NORMALIZZATO (ACN)*: l'importanza di ogni AMICO in COMUNE è inversamente proporzionale al logaritmo delle sue amicizie.

METRICHE BASATE SULLA TOPOLOGIA (1)

- *CENTRALITÀ DEGLI AMICI IN COMUNE (CENTR)*: l'importanza di ogni AMICO in COMUNE è pari alla sua centralità nella rete, calcolata tramite la sua *BETWEENNESS*.

Cos'è la BETWEENNESS?

Determina il livello di influenza di ogni nodo nell'essere un “punto di passaggio” nella rete (calcolando il numero di volte in cui si trova sul cammino più breve fra due nodi).

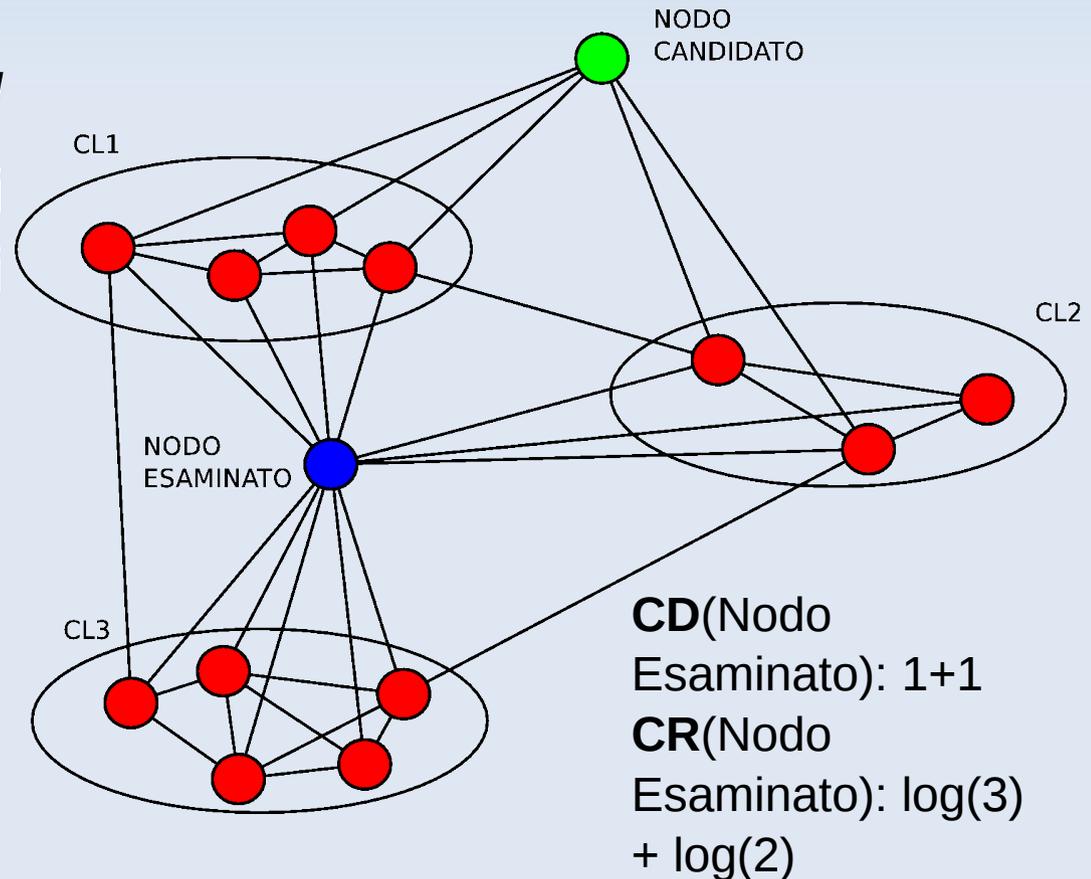


METRICHE BASATE SULLA TOPOLOGIA (2)

Cos'è un **CLUSTER**?

È un raggruppamento logico dei nodi all'interno di una rete, in cui sono riuniti quelli hanno un'alta interazione tra loro.

- *CAMMINI DISGIUNTI*
(CD): suggerisce i candidati raggiunti dal maggior numero di CLUSTER.
- *CAMMINI RAGGRUPPATI*
(CR): l'importanza di ogni CLUSTER è proporzionale al logaritmo del numero di AMICI in COMUNE al suo interno



IMPLEMENTAZIONE

Sono stati creati quattro Moduli Implementativi con lo scopo di garantire:

- La costruzione della PERSONAL NETWORK della PERSONA ESAMINATA
- La raccolta dei CANDIDATI da SUGGERIRE
- Il calcolo dei RANKING per ogni metrica
- La presentazione dei SUGGERIMENTI

VALUTAZIONE

È stata effettuata su un CAMPIONE di 23 utenti con un Profilo FACEBOOK attivo, a cui sono stati inviati via e-mail:

- Un file HTML con la lista dei CONTATTI SUGGERITI (Unione dei primi 10 generati per ogni metrica).
- Un file XLS consistente in una tabella-RESOCONTO da completare rispondendo a tre semplici domande per ogni contatto:
 - Conosci questa Persona?
 - Aggiungerai questa Persona?
 - Il suggerimento è interessante?

RISULTATI (1)

CORRELAZIONE TRA METRICHE

| RANK | AC | ACN | CENTR | CD | CR |
|-------|-------|-------|-------|-------|-------|
| AC | 1,000 | 0,631 | 0,424 | 0,690 | 0,999 |
| ACN | | 1,000 | 0,185 | 0,444 | 0,630 |
| CENTR | | | 1,000 | 0,320 | 0,424 |
| CD | | | | 1,000 | 0,718 |
| CR | | | | | 1,000 |

- Alta correlazione tra AC e CR (0,999)
- Poca invece (0,185) tra ACN e CENTR.
- Discreta tra CD e CR.

| VALORI | AC | ACN | CENTR | CD | CR |
|--------|-------|-------|-------|-------|-------|
| AC | 1,000 | 0,992 | 0,596 | 0,592 | 0,907 |
| ACN | | 1,000 | 0,583 | 0,583 | 0,897 |
| CENTR | | | 1,000 | 0,399 | 0,560 |
| CD | | | | 1,000 | 0,818 |
| CR | | | | | 1,000 |

- CENTR ha una discreta correlazione con tutte le altre metriche.
- AC e ACN sono molto correlate tra di loro.

RISULTATI (2)

SOVRAPPOSIZIONE METRICHE TRA I PRIMI 10 SUGGERIMENTI

1) La metrica CD ha poca sovrapposizione con le altre (2,913 max).

| | AC | ACN | CENTR | CD | CR |
|-------|--------|--------|--------|--------|--------|
| AC | 10,000 | 8,565 | 3,391 | 1,130 | 3,739 |
| ACN | | 10,000 | 3,826 | 1,478 | 3,478 |
| CENTR | | | 10,000 | 2,913 | 2,522 |
| CD | | | | 10,000 | 2,652 |
| CR | | | | | 10,000 |

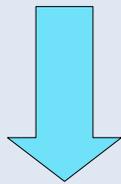
2) Tra AC e ACN c'è una grossa sovrapposizione tra i suggerimenti medi prodotti (8,565).

Si vuole trovare un modo per trovare suggerimenti diversi che AC non ottiene.

RISULTATI (3)

VALUTAZIONE dei SUGGERIMENTI

- Lo scostamento dei valori, soprattutto sulla risposta ad A, induce l'esecuzione di test statistici più precisi (POST-HOC) sulla variabilità dei dati.



- Per le risposte a A) le differenze statisticamente significative ottenute sono solo in presenza di CD.

| MEDIE - VARIANZA | A) Conosci? | B) Aggiungerai? | C) Interessante? |
|---------------------|----------------|-----------------|------------------|
| AC | 5,5 – 4,9 var. | 2,4 – 2,8 var. | 2,4 – 3,0 var. |
| ACN | 5,3 – 4,3 var. | 2,4 – 1,9 var. | 2,4 – 3,2 var. |
| CENTR | 4,3 – 3,4 var. | 2,3 – 2,5 var. | 2,3 – 2,8 var. |
| CD | 2,5 – 2,1 var. | 1,6 – 1,2 var. | 2,0 – 2,8 var. |
| CR | 4,2 – 3,1 var. | 2,3 – 2,0 var. | 2,4 – 2,1 var. |

- Nelle altre due risposte, il test non ha trovato significatività. Le metriche tendono dunque a comportarsi in modo simile.

RISULTATI (4)

SUGGERIMENTI INTERESSANTI

- AC è la metrica più semplice e non ha performance peggiori delle altre.
- Quali metriche hanno trovato più amici “interessanti” non suggeriti da AC?
 - I risultati ottenuti:
 - 0 interessanti per ACN (su una media di 1,5)
 - 1,3 interessanti per CENTR (su una media di 5,4)
 - 0,9 interessanti per CD (su una media di 6,2)
 - 0,6 interessanti per CR (su una media di 5,3)
- CD tende a trovare amici interessanti che nel ranking di AC sono molto spesso posizionati lontano dai primi dieci della graduatoria.

CONCLUSIONI

- Definizione di un metodo per un Sistema di Raccomandazione di contatti:
 - Automatico
 - Modulare
 - Rapido
- Ciascun sotto-problema:
 - È stato studiato dal punto di vista teorico
 - È stato affrontato progettando opportuni moduli software in grado di implementare le considerazioni teoriche

SVILUPPI FUTURI

- Studiare e caratterizzare nuove tecniche per:
 - sfruttare l'uso combinato delle metriche, differenziando in classi gli utenti da esaminare:
 - Persone
 - Aziende o Enti
 - introdurre, come Requisito indicato dall'utente, la preferenza sul grado di conoscenza dei contatti.