Il ruolo crescente che gioca la ***Social Network Analysis*** nella risoluzione dei crimini non solo ha aumentato la domanda di strumenti tecnologici, ma ha offerto anche opportunità illimitate alle ***nuove strategie di analisi*** precedentemente non disponibili. Vengono realizzati software che, da una parte rendano possibile la sua applicazione su una ***gran mole di dati*** e dall'altra, consentano di utilizzare ***nuove tipologie di misurazione*** dei dati, non ancora applicate nell'analisi di Intelligence, ma che sono estremamente efficaci nello svelare i ***modelli strutturali sottostanti le reti criminali*** e nel fornire quindi dei validi suggerimenti per la soluzione dei casi.

di Sara Boschin

SOLUZIONI INNOVATIVE APPLICATE ALL'ANALISI INVESTIGATIVA RIVOLTA ALLO STUDIO DELLE RETI CRIMINALI

Sistemi & Automazione

Sara BOSCHIN, analista e trainer presso Sistemi & Automazione dal 2012, affiancando gli utenti in molte attività di Polizia Giudiziaria, si è specializzata nell'analisi del traffico telefonico. Ha approfondito le sue conoscenze attraverso lo studio delle Reti Sociali, in particolare applicato a fenomeni criminali, dando così un notevole contributo alla ricerca di nuove ipotesi investigative e di soluzioni efficaci.

1. Introduzione

La criminalità organizzata è un sistema sociale complesso, con dinamiche e comportamenti tipici, che sfrutta le connessioni sociali tra i propri affiliati. Per comprenderla meglio e attivare strategie di prevenzione e contrasto alle sue attività, vengono condotti degli studi volti a scoprire schemi regolari della struttura e del flusso delle informazioni e quindi ad individuare dei "modelli" di analisi che consentano di identificare delle reti criminali con caratteristiche comuni.
Questo tipo di analisi richiede innanzitutto la capacità di integrare dati provenienti da molteplici eventi criminali e da più fonti. Una sovrabbondanza di dati che può produrre poco valore. Si hanno cioè grandi volumi di "dati grezzi" raccolti da più fonti: registrazioni telefoniche, conti bancari e transazioni, vendite di veicoli, relazioni di sorveglianza, solo per citarne alcune, però mancano sofisticati strumenti e tecniche di analisi per utilizzarli in modo efficace ed efficiente.
Cerchiamo di capire le metodologie attualmente impiegate in tale contesto e i nuovi orientamenti che stanno emergendo rispetto a questo tipo di esigenza.

2. Gli approcci metodologici e i modelli di analisi

Attualmente, lo studio delle reti in ambito investigativo, come ad esempio quello sulle reti criminali (terrorismo, narcotraffico, frode, bande criminali correlate, furto d'armi e altro), viene condotto principalmente con strumenti di analisi che presuppongono un approccio relazionale e visuale ai dati raccolti, chiamato *Link Analysis*.
Questo metodo, sviluppato da ANACAPA[1] negli anni '60, consente di rappresentare le informazioni tramite diagrammi, usando **Entità** e **Collegamenti**. I vantaggi sono notevoli:

- facilita la comprensione delle informazioni, grazie alla visualizzazione grafica delle entità coinvolte e dei legami fra queste, altrimenti non facilmente individuabili;
- semplifica scenari articolati e complessi;
- permette l'analisi temporale di catene logico-temporali, utile se si pone l'enfasi sulla sequenzialità degli eventi;
- svela collegamenti nascosti;
- agevola la sinergia tra fonti. Grazie alla presenza di Layer semantici e algoritmi che normalizzano i dati, è possibile collegare automaticamente le informazioni, indipendentemente dal sistema e dalla fonte di provenienza (Banca Dati privata, pubblica, Internet, tabulati telefonici e/o telematici, estratti conto bancari, ecc.).

A questo approccio si associa la metodologia di analisi delle relazioni sociali, nota come *Social Network Analysis* (SNA) o analisi delle reti sociali, sviluppatasi a partire dai contributi di Jacob Levi Moreno[2], fondatore della sociometria.
Nella teoria delle reti sociali, la società è studiata come una rete di relazioni fra gli individui che ne fanno parte. Le interazioni fra i vari attori modificano reciprocamente i loro comportamenti. Lo scopo principale dell'analisi di *network* è quello di individuare e analizzare i **legami** (*ties*) tra gli **individui** (*nodes*). Per comprendere le reti e gli elementi che ne fanno parte e quindi avere una vista sui ruoli e sui raggruppamenti, si ricerca la centralità di un'entità all'interno della rete, misurandone la posizione.

1 Anacapa Sciences, Inc. è una società con sede in California. Dal 1971 sviluppa le tecniche ed i metodi per l'analisi di intelligence che sono attualmente impiegati in tutto il mondo, organizzando anche corsi di formazione su questa tematica.

2 Jacob Levi Moreno è stato uno psichiatra statunitense di inizio Novecento, considerato il fondatore della "microsociologia", ovvero della osservazione microscopica dei fenomeni associativi e della "sociometria", la misurazione-rilevazione delle relazioni esistenti in un gruppo dato o in una comunità. La sua convinzione fu che il gruppo costituisca l'atomo funzionale delle dinamiche sociali, e che mescolandosi con altri gruppi formi strutture sempre più complesse.

Tale attività di mappatura aiuta a comprendere il comportamento globale di una rete e la sua capacità di raggiungere gli obiettivi chiave e consente di valutare le caratteristiche dei singoli nodi o delle strutture che questi creano mediante le relazioni, caratteristiche molto spesso non immediatamente riconoscibili. Un esempio può essere l'individuazione di una più piccola sotto rete operante all'interno di una rete più grande. L'applicazione di questa metodologia all'analisi delle organizzazioni criminali ha come presupposto il fatto che i crimini sono svolti principalmente da più soggetti che, collaborando, possono formare gruppi e giocare ruoli diversi. Risulta fondamentale capire chi conosce chi e chi fa affari con chi.

In una rete di traffico di stupefacenti, per fare un esempio, i diversi gruppi possono essere responsabili della gestione della fornitura di droga, della distribuzione, della vendita, del contrabbando e del riciclaggio di denaro. In ogni gruppo ci può essere un leader che comanda e fornisce meccanismi di direzione agli altri, così come i *gatekeepers* che assicurano che le informazioni e le merci circolino efficacemente tra i gruppi.[3]
Le metriche della *social network analysis*, sono:

- Il *Degree*, che consente di trovare le entità più attive all'interno della rete, considerando quelle che hanno il maggior numero di collegamenti diretti rispetto alle entità presenti.
- La *Closeness*, utile per individuare l'entità che raggiunge gli altri componenti della rete in maniera più rapida.[4]
- La *Betweenneess* che individua l'entità che svolge il ruolo strategico di veicolare e controllare il flusso di informazioni in zone differenti della rete. Sono i cosiddetti *gatekeepers*, gli "intermediari", di cui si è fatto cenno in precedenza.
- l'*Eigenvector* per determinare l'entità con un 'forte capitale sociale', in quanto connessa con persone chiave, molto attive, in grado di esercitare una grande influenza nelle scelte degli individui.
- il *K-Core*[5] che identifica le comunità, cioè dei piccoli gruppi di soggetti fortemente coesi all'interno di una rete più vasta.

Questi algoritmi si applicano al *network* per individuare dei *pattern*, cioè delle caratteristiche tipiche, modalità di comportamento specifiche, che possono essere contestualizzate alle reti criminali.[6] È il caso, ad esempio, del *K-Core*, che consente l'individuazione del *pattern* conosciuto col nome di "**effetto piccolo mondo**": questo identifica una rete avente una elevata connettività. Al suo interno tutti i nodi vantano un numero elevato di link che, di fatto, accorciano le distanze fra di loro all'interno del sistema.
Un altro *pattern* interessante è quello delle reti "**scale-free**", caratterizzate dalla presenza di alcuni soggetti, la cui eliminazione è in grado di causare anche l'interruzione del rapporto tra due realtà criminali separate. Ovviamente, l'efficacia della rimozione di un individuo o di un insieme di individui in una rete dipende largamente dal "fattore unicità" e, di conseguenza, dai fattori "sostituibilità" ed "equivalenza di ruolo" di altri nodi presenti nella rete.[7] Quindi, conseguentemente alla rimozione, casuale o mirata che sia, occorre considerare la *resilience* della rete, cioè la sua capacità rigenerativa, utile per pianificare efficaci azioni di contrasto all'attività delinquenziale.

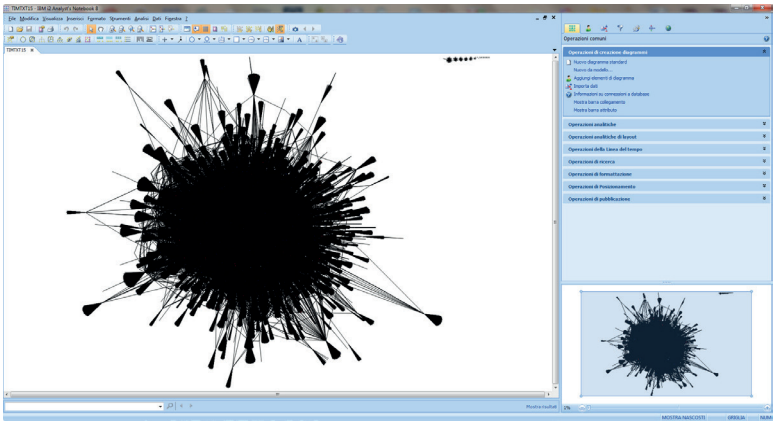
Nell'analisi dei network sono da considerare anche la *Link Direction* (direzione del collegamento) e la *Link Weightings* (pesatura del collegamento). La prima risulta fondamentale quando si analizza del traffico telefonico o delle transazioni bancarie, per comprendere la direzione dei flussi; la seconda invece è utilizzata per attribuire un'importanza diversa ai legami tra le entità come, ad esempio, gli affiliati di un'organizzazione criminale. Pesando le relazioni fra le persone e i percorsi costruiti da queste all'interno della rete, si può comprendere meglio il ruolo che ciascuno ricopre al suo interno.

Un interessante ambito di applicazione della metodologia SNA è l'**analisi del traffico telefonico**. Questa attività è, nella maggior parte dei casi, caratterizzata da una gran mole di dati. Spesso il traffico che viene analizzato comprende un arco temporale molto ampio, difficile da esaminare anche con l'approccio visuale.

***Figura 1**: l'immagine mostra un esempio di analisi di questo tipo, condotta utilizzando il visualizzatore **Analyst's Notebook**. È stato inviato su chart il traffico telefonico di alcune utenze di particolare interesse investigativo, relativo a diversi mesi.*

Applicare le metriche della SNA ad un grafico come quello rappresentato nell'immagine, è di fatto un'operazione poco utile, data la scarsa comprensione dei dati contenuti.

L'unica soluzione in questo caso, sarebbe quella di scorporare il traffico su più diagrammi, ad esempio per periodi temporali ben circoscritti, ed applicare poi su ciascuno di essi le misurazioni, rinviando ad un momento successivo la comparazione dei vari grafici.



3 *Criminal Network Analysis and Visualization*. Jennifer Xu, Hsinchun Chen.

4 *Social network analysis in an operational environment: Defining the utility of a network approach for crime analysis using the Richmond City Police Department as a case study*. Jennifer A. Johnson and John David Reitzel.

5 *Generating Graphs with Predefined k-Core Structure*. Michael Baur, Marco Gaertler, Robert Gorke, Marcus Krug, e Dorothea Wagner.

6 Su tali argomenti, Sistemi & Automazione realizza dei corsi di formazione specifici, volti all'apprendimento dell'applicazione pratica di tali misurazioni alle varie realtà investigative.

7 Per ulteriori approfondimenti su questo tipo di ricerca, si rimanda a *International Journal of Intelligence and Counter Intelligence*, Volume 5, No. 3. Articolo: *Network Vulnerabilities and Strategic Intelligence in Law Enforcement*. Malcolm K. Sparrow.

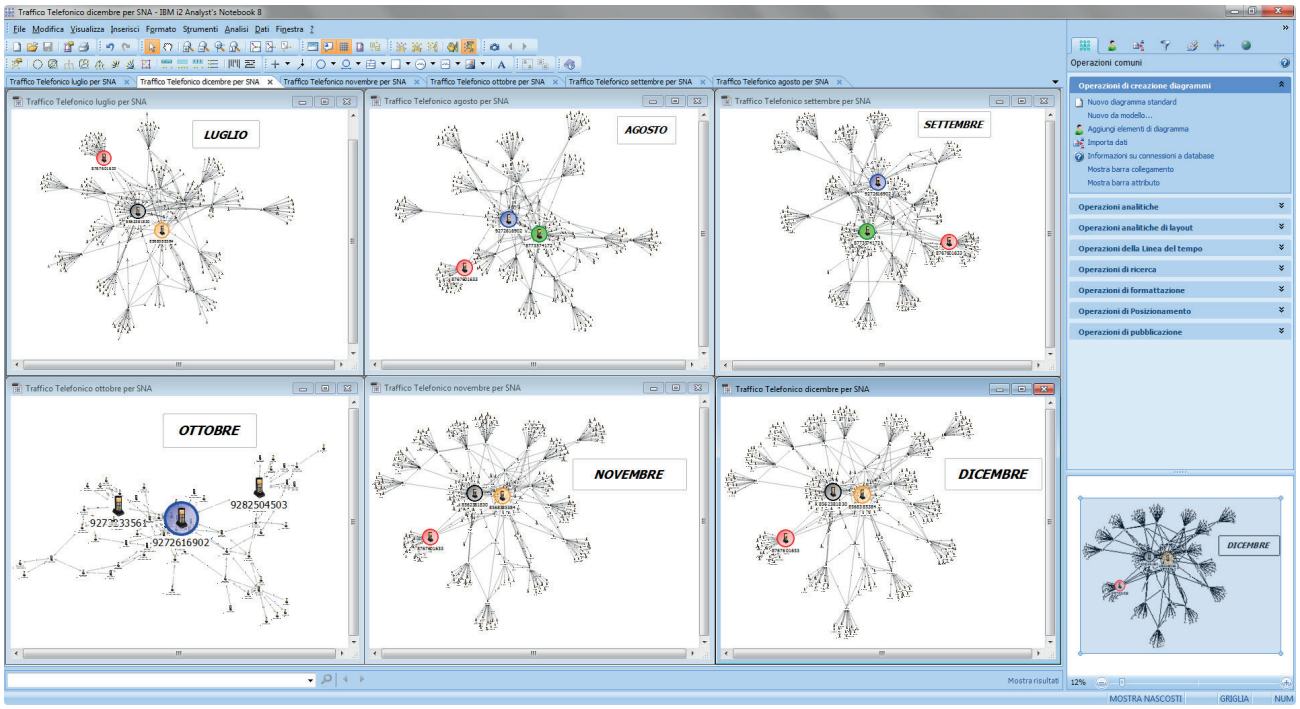


***Figura 2**: l'immagine mostra la suddivisione su più diagrammi di un traffico telefonico semestrale, di particolare interesse investigativo.*

In questo specifico caso, sono stati utilizzati due prodotti informatici: il visualizzatore *Analyst's Notebook* e il software per la gestione dei dati di traffico telefonico **Tetras 2 HPG**, fornito da Sistemi & Automazione s.r.l., società leader in Italia nella realizzazione di soluzioni e prodotti per l'analisi delle informazioni e l'Intelligence. Dopo aver suddiviso il traffico telefonico su più diagrammi, si è proceduto all'analisi dei dati. È stata calcolata la *betweenneess* per ogni mese e sono state considerate le prime tre utenze che risultavano avere i valori più elevati, messe in evidenza con una cornice colorata.
Dalla comparazione dei vari diagrammi, è risultato che l'utenza con la cornice nera aveva i valori di *betweenneess* più alti, ma era presente solo in tre dei sei mesi, quella con la cornice rossa risultava sempre al terzo posto nell'elenco dei valori derivanti dal calcolo ed era quasi sempre presente; infine, l'utenza con la cornice blu, aveva valori sempre alti ma compariva solo in tre dei sei mesi.
Approfondendo l'analisi sui collegamenti, è emerso anche che le utenze con la cornice rossa e nera erano sempre direttamente collegate, mostrando quindi un canale diretto di comunicazione. In sintesi, dal calcolo della centralità, l'utenza con la cornice rossa appariva ricoprire fra tutte il ruolo più strategico.
Tale attività può aiutare l'analista a conoscere l'effettiva importanza degli attori di una rete, le evoluzioni nel tempo, i cambi di ruolo e ad individuare le persone chiave per prendere opportune decisioni.
Da un punto di vista pratico, l'attività di analisi, condotta con questo procedimento, è caratterizzata da un notevole dispendio di tempo, oltre che dalla inevitabile perdita di alcuni legami fra le entità che possono avere una certa importanza. L'analista, dovendo evidenziare solo alcune parti della rete, data la grande mole di dati che deve gestire, deve di fatto rinunciare ad un'analisi preliminare degli stessi. Vi è quindi un oggettivo limite quantitativo.

3. Gli sviluppi tecnologici

L'esigenza di analizzare reti sempre più complesse, per tipologia di dati e dimensioni, ha condotto la ricerca, sia in ambito aziendale che universitario, ad investire notevoli risorse nello studio di algoritmi per il calcolo degli indici sempre più efficienti in termini di vincoli di tempo e di memoria.
In particolare, tali studi sono volti al superamento dei seguenti limiti oggettivi:

- la quantità di dati da analizzare;
- l'ambiguità delle reti;
- l'evoluzione nel tempo delle reti;
- la non attendibilità dei dati stessi.

Per far fronte ad un'utenza sempre più esigente nel campo dell'analisi delle reti applicata al mondo del crimine, anche Sistemi & Automazione, in collaborazione con il Dipartimento di Ingegneria Informatica, Automatica e Gestionale dell'Università di Roma "La Sapienza", ha risposto a questa nuova sfida avviando un progetto di ricerca e sviluppo di un software capace di fare dell'analisi delle reti sociali su grandi volumi di dati il punto di forza. Il presupposto fondamentale alla base del progetto è stato la definizione e predisposizione di opportune architetture hardware, integrabili con altre applicazioni che, tramite meccanismi di parallelizzazione e di ottimizzazione delle risorse, potessero consentire il raggiungimento di elevati miglioramenti in termini di efficienza computazionale. Il primo passo per raggiungere tale obiettivo è stato l'individuazione e la raccolta dei requisiti necessari alla realizzazione di un prodotto davvero innovativo, avvenuta anche grazie all'esperienza che le risorse aziendali hanno maturato a fianco della tipologia di utenti di cui abbiamo parlato in precedenza.
In primo luogo, è stata posta grande enfasi allo studio di nuovi algoritmi per il calcolo degli indicatori in ambito *BIG DATA*, nell'ottica dell'analisi di grafi complessi. La novità in questo caso sta nel fornire in tempi accettabili i risultati derivanti dall'applicazione delle metriche attualmente disponibili, ad una rete di grandi dimensioni.

I risultati ottenuti però devono essere anche leggibili. Un sistema innovativo deve fornire una visualizzazione efficace degli stessi e, ad esempio, rendere individuabili i nodi con alti valori di *betweenness*, *closeness*, ecc. anche in grafici di grandi dimensioni e in contesti applicativi differenti (marketing, antifrode, medicina, organizzazioni militari).
Classificare una rete come appartenente ad una specifica tipologia di fenomeno, è un'attività molto utile per l'analista in termini predittivi, in quanto permette di anticipare il comportamento, definire e attuare azioni di contrasto, volte a disarticolare la rete stessa.
Appare molto vantaggioso poter disporre di nuove metriche di analisi avanzata, mai applicate finora all'analisi di Intelligence, ma estremamente utili per il raggiungimento di tale obiettivo. Rispetto a questa esigenza, è stata avviata un'indagine direttamente sull'utenza destinataria del software (principalmente analisti del settore investigativo), finalizzata alla raccolta di idee e suggerimenti, utili allo sviluppo di questi nuovi strumenti di analisi. Al termine, sono stati indentificati alcuni ***pattern* ricorrenti**, cioè degli **schemi di comportamento** che si ripetono secondo determinate caratteristiche o variabili (il tempo, la quantità di legami, i valori inclusi nei collegamenti, la direzione degli stessi), sui quali sviluppare i *pattern* matematici che genereranno uno o più algoritmi da applicare a contesti molto ricchi di dati.

Alcuni dei *pattern* considerati nel progetto, sono già noti a chi lavora in questo settore. Altri invece sono emersi nella fase di studio e raccolta dei requisiti. Tra quelli noti, in particolare a chi si occupa di analisi del traffico telefonico per contrastare la criminalità organizzata, c'è il cosiddetto "citofono" o meglio "citofoni", con cui ci si riferisce a quelle utenze dedicate a una relazione telefonica esclusiva che si contattano reciprocamente, senza avere relazioni con altre utenze.
È una modalità operativa utilizzata principalmente dai latitanti e viene rappresentata con una coppia di utenze collegate fra loro e separate dal resto della rete. Questo tipo di rappresentazione non basta per individuare il *pattern* e si ricorre all'identificazione di altre caratteristiche, fra cui:

- **la direzione dei collegamenti**: spesso univoca, perché di solito è il latitante che chiama;
- **l'impegno della cella**: ci dovrebbe essere una omogeneità temporale nell'aggancio delle celle. Il latitante rimane nascosto per un periodo temporale più o meno lungo in una determinata zona. Poi si sposta;
- **la fascia oraria**: si potrebbe creare un periodo temporale di preferenza in cui effettuare le chiamate, ad esempio in fasce orarie di punta, per mimetizzarsi in mezzo a grandi moli di dati.

Sempre nell'ambito dell'analisi del traffico telefonico, un altro *pattern* interessante e già conosciuto è il "Palo", che coincide con l'utenza che impegna la stessa cella o il gruppo di celle della medesima zona e negli stessi orari di altre utenze attenzionate. Nel concreto, è il classico caso del rapinatore che funge da palo, appunto, perché rimane nelle vicinanze della zona nella quale avviene la rapina, per avvisare i complici in caso di imprevisto. L'analisi, in questo caso, si concentra sulla rappresentazione temporale delle telefonate e sul legame con le celle.

Un altro nuovo *pattern* è il "*fork*", caratterizzato dalla presenza di un'entità con quattro collegamenti, di cui uno in entrata e tre in uscita. È uno schema ricorrente nei casi di sequestro di persona, quando si ha un'utenza che riceve una chiamata (collegamento in entrata) e che poi ne effettua consecutivamente tre (collegamenti in uscita) con un intervallo temporale particolarmente ravvicinato alla prima chiamata in ingresso. Un altro esempio di *fork* può essere rappresentato da un conto corrente che riceve un bonifico (collegamento in entrata) dall'importo consistente ed effettua immediatamente dopo tre transazioni (collegamenti in uscita), in un arco temporale molto ristretto.
Rispetto alle caratteristiche più prettamente "tecniche" che il software deve possedere, è emersa la necessità di non avere dei costi computazionali troppo elevati a livello di memoria e di tempo e questo sia per la tipologia di attività svolta attraverso questi strumenti, sia per il tipo di strumentazione hardware utilizzata da chi opera in questo settore che, nella maggior parte dei casi, non è di livello elevato. Infine, il prodotto deve risultare fruibile all'utente finale: un software, anche se potente e ricco di funzioni, è infatti tanto più apprezzato da chi lo utilizza, quanto più risulta intuitivo e di semplice comprensione nell'uso.

4. Conclusioni

Nell'attuale contesto investigativo, le esigenze di analisi rivolte allo studio delle caratteristiche tipiche delle reti criminali impongono di considerare metodologie e strumenti che facciano del calcolo statistico e computazionale il loro punto di forza.
L'ambito in cui ci si muove infatti, è caratterizzato da una grande quantità di informazioni provenienti da più fonti e dalla necessità di districarsi fra questi dati ricavandone suggerimenti utili per prevedere i comportamenti delle organizzazioni criminali e per attuare ipotesi di contrasto alla loro azione delinquenziale. In tal senso la SNA, come metodologia di analisi, offre già a chi si occupa di queste realtà un prezioso apporto in termini di comprensione degli scenari. Sempre più spesso infatti, viene adottata dalla comunità di investigatori perché aiuta a massimizzare il valore delle informazioni raccolte.

Il limite attuale è di carattere quantitativo, poiché risulta difficile applicare le metriche a diagrammi molto complessi, senza perdere di vista alcuni importanti aspetti. D'altro canto, l'identificazione di nuovi *pattern* sui quali sviluppare nuovi algoritmi matematici, potrebbe essere utile per far emergere delle prospettive di analisi finora non sufficientemente considerate.
Con il supporto di una tecnologia robusta, l'analisi delle reti sociali può diventare uno strumento ricco e affidabile nel tempo e i risultati prodotti risultare fruibili all'interno di qualsiasi ambiente operativo di polizia.
La spinta attuale muove quindi verso il suo potenziamento, mediante lo studio di nuove architetture hardware rispondenti ai criteri suggeriti. Il progetto di Sistemi & Automazione costituirà una grande innovazione dal punto di vista dell'analisi delle reti sociali, non solo per la presenza di sofisticati algoritmi per il calcolo delle misure SNA, sia di quelle esistenti che delle nuove, applicate su larga scala, ma anche per la possibilità di far girare il prodotto in ambienti *enterprise*, dove solitamente non sono disponibili hardware e software specifici per la ricerca ed il calcolo scientifico. ©