



# NUOVE FRONTIERE DELL'ANALISI VISUALE PER L'ESPLORAZIONE DI RETI SOCIALI

**MARCO ANGELINI**

*L'uso pervasivo delle reti sociali per scopi e contenuti eterogenei ha ingenerato l'interesse per lo studio delle informazioni, con la difficoltà di ricorrere ai normali strumenti d'indagine. Laddove l'analisi automatica si dimostri insufficiente e quella umana risulti più puntuale ma meno scalabile, la combinazione delle stesse tramite Visual Analytics fornisce un metodo per l'esame di questi dati. L'articolo espone una panoramica sulle nuove frontiere dell'analisi visuale applicata alle reti sociali, in termini di opportunità, risultati e sfide da risolvere.*

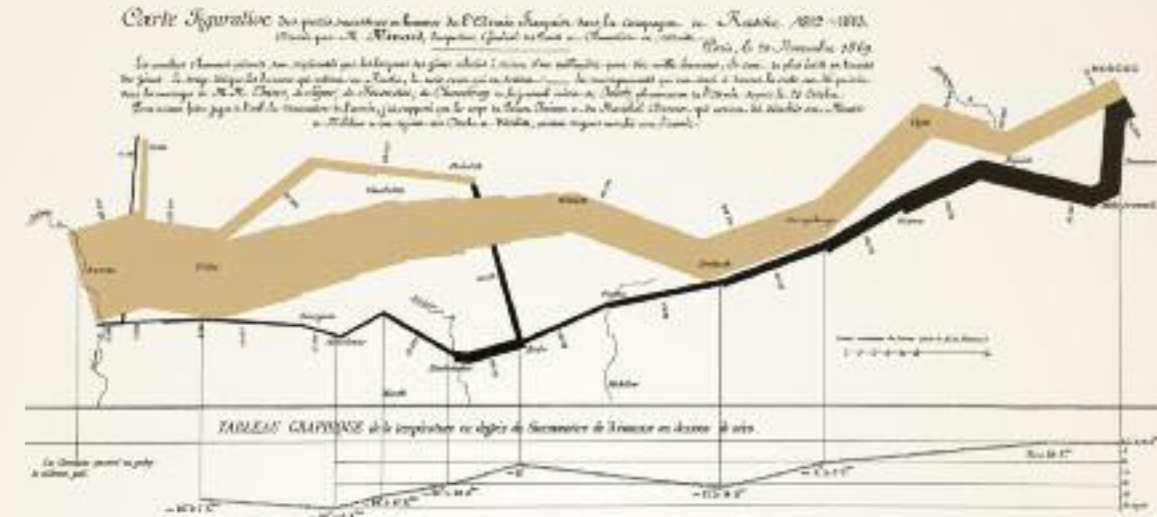
Il ritmo odierno di produzione di dati, siano essi in relazione ai nuovi dispositivi smart ovvero derivanti dall'analisi di fenomeni complessi, continua a imprimere una spinta molto forte al progresso tecnologico, culturale, scientifico e umano delle moderne società. In tale ambito, la capacità di collezionare questi dati anche a livello del singolo individuo (biomedici, relativi alle preferenze di acquisto o alla geo-localizzazione) e il supporto tecnologico alla loro condivisione hanno portato dapprima alla nascita dei social network (reti sociali) e successivamente hanno conferito loro sempre maggiore importanza.

L'aggregatore di dati statistici online <statista.com> stima che a oggi, nelle varie forme in cui esistono le reti sociali, più di 2,60 miliardi di persone ne utilizzino almeno una, con la proiezione di più di tre miliardi entro il 2021. Per rafforzare il concetto, nella figura della pagina seguente è riportata un'istantanea dello stato di diffusione del solo Facebook. A fronte di indubbi vantaggi e opportunità che offrono ai propri utenti, e più in generale allo studio delle dinamiche collettive – si pensi a esperimenti come la *Bucket-challenge*, oppure alle campagne di sensibilizzazione – le reti sociali introducono anche elementi controversi per alcune criticità, come la gestione della privacy, la diffusione di fake news, e il loro utilizzo per campagne di *phishing*.



Visualizzazione del livello di diffusione del social network Facebook all'anno 2017.

Sorgono, pertanto, il bisogno e l'opportunità di analizzare questi dati, con il necessario ricorso a strumenti automatici, che consentano di raggiungere interessanti risultati, anche se limitati dalle capacità algoritmiche degli attuali dispositivi di calcolo e dalle caratteristiche di alta cardinalità, eterogeneità e velocità di cambiamento dei dati prodotti in una rete sociale. Ancora più sfidante come tema è rendere fruibili tali risultati per l'analisi a un essere umano, spesso elemento finale di ogni processo decisionale. A tal proposito, fin dal 1800 la rappresentazione dell'informazione è stata usata come strumento di studio e spiegazione di fenomeni complessi: si pensi alle visualizzazioni statiche create, durante la sua carriera, dall'ingegnere civile francese Charles Joseph Minard (1781-1870), di cui la più nota, realizzata nel 1869 per analizzare la Campagna di Russia di Napoleone (1812-1813), è riportata nella figura della pagina seguente: in forma compatta, sono riprodotte informazioni spazio-temporali come l'andamento del numero di truppe e della temperatura durante la Campagna, le principali battaglie, i guadi e gli attraversamenti geografici relativi all'avanzata verso Mosca (in marrone) e la ritirata verso Parigi (in nero). La rappresentazione, creata staticamente e senza l'aiuto dei calcolatori, è un ottimo esempio di come essa possa aiutare l'essere umano nella comprensione di fenomeni complessi caratterizzati da dati multidimensionali.



Mappa di Charles Joseph Minard della disastrosa Campagna russa di Napoleone del 1812.

La visualizzazione è il primo esempio noto della tecnica del *sankey* in due dimensioni e rappresenta sei tipi di dato: numero delle truppe di Napoleone, distanza, temperatura, altitudine e longitudine, senso di marcia e posizione relativa a date specifiche.

Tornando allo scenario attuale, il coinvolgimento dell'utente e la sua interazione sono fondamentali in tutti i processi di esplorazione dei dati e, in particolare, per quelli derivanti dalle reti sociali, qualora l'obiettivo dell'analisi non sia completamente definito a priori e si voglia mettere l'analista nella condizione di poterlo fare. Per ottenere queste capacità è stata creata, prima, la moderna *Data Visualization* (rappresentazione visuale dei dati supportata dai calcolatori) e, più recentemente, la disciplina di *Visual Analytics* (analisi visuale).

L'analisi visuale è «la scienza del ragionamento analitico facilitata da rappresentazioni visuali interattive»<sup>1</sup>. Può attaccare problemi che risultano intrattabili dal punto di vista puramente computazionale tramite una stretta collaborazione tra l'uomo e la computazione automatica. Questa collaborazione è resa possibile visualizzando i risultati della computazione in modo da renderli fruibili all'utente, e trasformando le interazioni dell'utente in parametrizzazioni di modelli e filtri sui dati per il lancio di nuovi processi automatici, in un ciclo che si ripete e a ogni iterazione porta all'estrazione di nuova conoscenza. L'analisi visuale avanza gli sviluppi scientifici e tecnologici nel ragionamento analitico, nell'interazione, nelle trasformazioni e nelle rappresentazioni dei dati per il calcolo e la visualizzazione.

1. COOK – THOMAS 2005.



Come agenda di ricerca, essa riunisce diverse comunità scientifiche, domini di interesse e tecniche: dall'informatica, l'*information visualization* (visualizzazione dell'informazione), le scienze cognitive e percettive, il design interattivo, la progettazione grafica e le scienze sociali. Anche al netto dello stato dell'arte delle soluzioni esistenti, tale analisi pone diverse sfide. Questo articolo si propone di delinearne le principali e fornire spunti per l'evoluzione delle soluzioni.

In merito all'analisi di una rete sociale, a prescindere dal tipo di rete e dai tipi di analisi<sup>2</sup>, un passo preliminare riguarda la collezione dell'informazione, o *mining*: la quantità di dati e informazioni su una rete sociale è molto elevata e sta crescendo a ritmi assai rapidi.

Nelle reti sociali possono essere trovate tutte le tipologie di dati: strutturati, semi-strutturati o non strutturati sotto forma di testi, video, immagini, collegamenti ipertestuali o metadati. Anche la sola selezione dei campioni di dati adatti che possano essere rappresentativi della rete sociale sotto analisi è una sfida. In aggiunta, molti di questi dati potrebbero non avere un buon livello di qualità, e non esiste una procedura definita per controllarlo. In tale ambito l'analisi visuale offre all'analista la capacità di esplorare la qualità dei dati, attraverso tecniche di analisi del testo i cui risultati vengono visualizzati in forma opportuna (ad esempio, tramite *wordcloud*), permettendo di identificarne la qualità ed eventualmente correggerla, di superare i limiti della pura computazione automatica ed eventualmente di utilizzare i dati ottenuti a più alta qualità per istruire processi di classificazione automatica che meglio li supportino.

Successivamente alla raccolta dei dati, una prima analisi d'interesse riguarda l'estrazione e l'analisi delle relazioni esistenti tra gli elementi della rete sociale (link). La corretta modellazione di questa informazione, a elevata cardinalità e di natura dinamica (relazioni possono essere create, cancellate o modificate nel tempo), ha un impatto sul processo di estrazione. La selezione di una rappresentazione visuale appropriata ha invece un impatto sull'analisi dei link, con paradigmi differenti che meglio si prestano a specifici tipi di relazioni e che spesso sono usati congiuntamente. La più comune resta quella tramite un grafo composto da nodi e archi, anche se la cardinalità dei dati analizzati la rende poco scalabile e, se usata nella sua forma base, di difficile interpretazione. Il numero di trasformazioni possibili aumenta con la complessità della rete sociale, con possibilità multiple di significato sia per i nodi che per gli archi.

2. GOSWAMI – KUMAR 2017, p. 4021.

Scegliere una visualizzazione appropriata ha un impatto significativo sulla qualità delle inferenze statistiche che possono essere ricavate. A tal proposito, le reti sociali multidimensionali sono definite come le reti in cui esistono diversi tipi di relazioni tra i nodi, che possono essere raggruppate in livelli, per tipo omogeneo di relazione. Una relazione può essere un link diretto di tipo sociale tra utenti (esempio: relazione di amicizia) oppure un collegamento semantico tra utenti e contenuti (esempio: autore di un post). Relazioni diverse hanno significati diversi, come collegamenti positivi e negativi ad 'amici' e 'nemici'. L'analisi visuale può gestire questo tipo di reti in forma stratificata, permettendo di selezionare o deselezionare i livelli di relazione scelti e tramite questo meccanismo procedere per raffinamenti successivi all'analisi di queste reti, distinguendo gli strati di interesse.

In aggiunta, spesso le analisi si concentrano su porzioni della rete sociale: difatti una rete sociale è costituita da un'enorme quantità di dati, non tutti necessari allo stesso momento. La loro elaborazione per la scoperta di sottografi di interesse è dispendiosa dal punto di vista computazionale, così come la capacità d'identificare pattern temporali nella creazione o modifica dei collegamenti. L'analisi visuale rende entrambe le attività meno pesanti mettendo l'analista in grado di identificare aree di interesse o pattern temporali di rilievo tramite la selezione manuale di porzioni della rete sociale (esempio: comunità oppure pattern d'interazione). Questa selezione può essere realizzata tramite rappresentazioni visuali interattive, il cui grado di astrazione può abbracciare diversi paradigmi contemporaneamente, come matrici di adiacenza per componenti fortemente connesse della rete legate tra loro con classici archi che richiamano il paradigma visuale nodo-arco ma lo arricchiscono nelle aree più fitte di interazioni.

L'analisi visuale basata sulla rappresentazione delle reti è molto diffusa. Tra i tipi di rete modellati assumono particolare rilievo quelle di *follower* (i nodi rappresentano utenti e gli archi relazioni di affiliazione) e quelle di diffusione (i nodi rappresentano contenuti e gli archi relazioni di diffusione come menzioni o risposte). Queste rappresentazioni sono spesso coadiuvate da analisi a vari livelli di aggregazione, come analisi delle parole chiave sui contenuti (messaggi, video), analisi di diffusione geografica dei contenuti e correlazione con le caratteristiche topologiche della rete sociale, ad esempio utilizzando i dati dei social media geo-tagati per arricchire le analisi sulle traiettorie di movimento dei veicoli, aiutando gli utenti a capire meglio la semantica del movimento. *Sentiment analysis*, *hate speech detection* e *topic modeling*, sia sui contenuti che sulle eventuali conversazioni, sono altri tipi di analisi che vengono condotte sulle reti sociali e attualmente supportate dall'analisi visuale, spesso però su collezioni statiche di dati (ottenute tramite campionamento) e con grado di accuratezza variabile.

È possibile trovare nella letteratura interessanti modelli evolutivi di visualizzazione delle reti sociali<sup>3</sup>. All'inizio, la ricerca si è concentrata sulla visualizzazione della loro topologia applicando le tecniche esistenti per analizzare le reti di follower nelle reti sociali. Solo in una fase successiva la ricerca ha prestato attenzione alla visualizzazione e all'analisi del processo di diffusione delle informazioni. Contributi alla visualizzazione di queste reti annoverano sforzi in direzioni diverse, tra cui l'analisi delle ego-network, delle comunità e dei topic. Di grande interesse è correlare questi modelli di rete derivati dall'analisi tra loro, in particolare la rete di diffusione con la rete di follower, al fine di descrivere in forma più approfondita le dinamiche di comunicazione su una rete sociale. L'analisi su come le informazioni siano diffuse attraverso la rete di follower consente agli utenti di comprendere meglio il processo di diffusione stesso e la modalità con cui i differenti attori vi partecipano. La ricerca e sviluppo di soluzioni visuali che permettano l'analisi congiunta di queste reti, consentendo all'utente di passare in modo fluido tra queste due rappresentazioni mentre conduce il suo processo di analisi (ad esempio selezionando insieme di utenti e relativi contenuti di cui studiare la diffusione su ulteriori insiemi di utenti), rendono più gestibile la computazione in base alle richieste dell'utente e, al contempo, aumentano e rendono più versatili le analisi applicabili.

Pur con tutti i vantaggi elencati e le casistiche in cui l'analisi visuale porta attualmente un vantaggio – quali maggiore comprensione dei fenomeni, capacità di usare il ragionamento umano per superare i limiti del puro approccio automatico, capacità di coinvolgere più a fondo l'utente nel processo di analisi – resta aperta una serie di aspetti che vengono elencati di seguito come sfide di ricerca e scenari futuri.

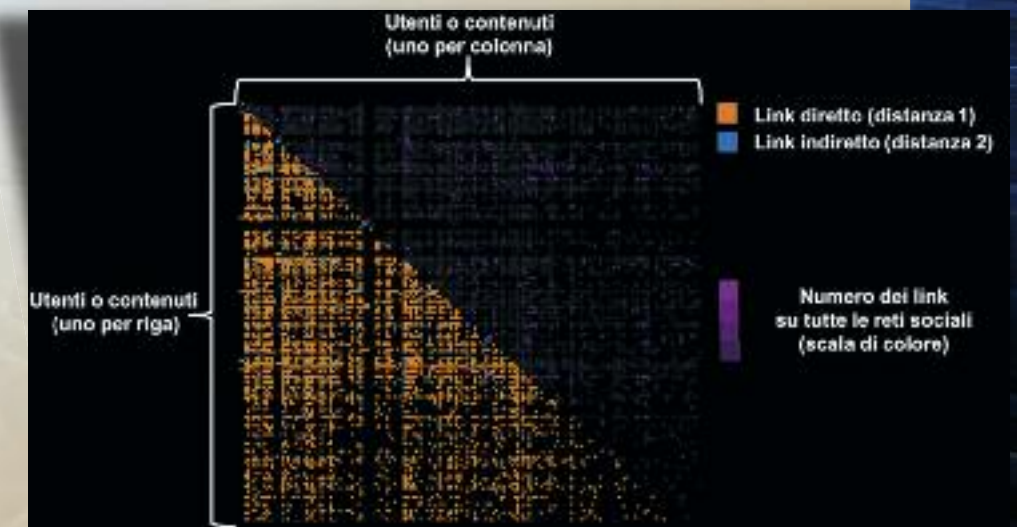
Un primo elemento di analisi è l'estrazione di comportamenti legati agli utenti di una rete sociale, dove un comportamento può essere modellato come un insieme strutturato di azioni semplici effettuate su di una singola rete sociale (ad esempio, la diffusione di un contenuto e i like a contenuti analoghi), caratterizzato ulteriormente da metriche riguardanti i contenuti stessi su cui l'utente interagisce e i pattern temporali (ripetizione periodica dello stesso insieme di azioni su contenuti simili). Questo tipo di analisi pone delle sfide sia per il livello di astrazione con cui rappresentare gli elementi di più alto livello rispetto alle interazioni di base, sia per il costo computazionale dovuto alle stesse analisi, spesso basate su algoritmi di *clustering* e classificazione, che potrebbe rendere lenta la produzione dei dati e quindi l'interazione dell'essere umano.

Una seconda sfida riguarda l'integrazione multi-rete di questi dati, con lo sviluppo di modelli che permettano di legare attività diverse allo stesso attore e di correlare elementi provenienti da reti differenti al fine di estrarre evidenze o nuova conoscenza. Ciò impone dei nuovi requisiti, come quello di mantenere l'esplorazione del contesto vicino ai dati di interesse, seguendo un principio di località: mantenere al

3. CHEN ET AL. 2017.

centro della rappresentazione i dati di interesse, arricchendo la visualizzazione nella parte periferica di elementi non selezionati ma a distanza minima da essi nella stessa rete sociale (ad esempio, tutti gli utenti a distanza di amicizia 1 dagli utenti di interesse) e a elementi direttamente connessi ma appartenenti ad altre reti sociali, permettendo di continuare l'analisi su reti sociali differenti da quella di partenza. Si assiste, quindi, alla necessità di ottenere contemporaneamente una visione d'insieme (*overview*), affiancata alla capacità di ottenere dettagli su elementi o reti specifiche tramite interazione. Questo requisito di esplorazione è comune anche ad altri campi della ricerca (si pensi al nuovo campo della *network medicine* o medicina di precisione, che lega elementi del genoma umano ai genotipi e fenotipi di malattie e a caratteristiche dei medicinali, dove ognuna di queste entità è rappresentata attraverso una rete di interazioni) e rappresenta uno scenario sfidante per le future applicazioni di analisi visuale. Caratteristiche fondamentali da considerare sul tema dovrebbero includere la capacità di fornire contemporaneamente un'analisi aggregata multi-rete, esplorabile tramite la selezione di dati sulle singole reti sociali d'interesse.

Un esempio è riportato nella figura in basso, dove tramite una matrice d'incidenza che lega tra loro utenti di più reti sociali (rappresentati come le righe e le colonne di questa matrice quadrata) è possibile, nella parte inferiore, rappresentare le loro interazioni dirette in arancione oppure indirette, ovvero con almeno un ulteriore attore o contenuto nel mezzo, in arancione su una singola rete. Nella parte superiore lo stesso sistema è utilizzato per




Esempio di rappresentazione di una comunità di utenti su più reti sociali. La matrice triangolare superiore rappresenta l'aggregazione delle interazioni su tutte le reti sociali considerate. La matrice triangolare inferiore è istanziabile su una singola rete sociale.

aggregare le possibili interazioni su un più ampio insieme di reti sociali. È così possibile, tramite l'esplorazione, analizzare le forme di interazione tra utenti più pronunciate nella parte superiore, e successivamente analizzare su quali reti queste forniscano il maggior contributo. Analogamente è possibile identificare anomalie rappresentate da utenti con un basso livello di interazioni su molte delle reti sociali considerate, ma alti picchi di attività su reti sociali specifiche. Chiaramente attività come *entity reconciliation* e disambiguazione dei contenuti rappresentano tuttora elementi di ricerca attiva.

Un'ulteriore sfida è costituita dal mantenere la velocità di analisi visuale dei dati al pari di quella del loro cambiamento nella rete sociale. Esistono computazioni e analisi da applicare a una rete sociale che per il numero di dati sotto esame e il loro rapido cambiamento non riescono a essere completate tempestivamente (ad esempio, l'estrazione di sottografi di interesse o l'analisi del cambiamento su una rete sociale), e richiedono comunque molto tempo prima di produrre un risultato da visualizzare. Ciò rende molto difficile il processo di analisi dei dati e impossibile la loro esplorazione per via dei lunghi tempi di attesa. A tal fine, metodi di analisi progressiva, basati sulla computazione incrementale di un insieme di dati che produca in tempi rapidi una stima del risultato dell'analisi, convergente nel tempo al risultato esatto, permette la produzione di visualizzazioni in tempi molto più rapidi, e all'utente l'analisi e l'esplorazione dei dati di interesse: di contro, introduce la necessità di gestire il grado di confidenza e incertezza dei risultati parziali prodotti e di fornire all'utente la loro corretta interpretazione. Ad esempio, in un processo di analisi visuale della diffusione di fake news, il sistema potrebbe rendere disponibile un set iniziale di nodi della rete sociale con maggiore livello di diffusione con un determinato indice di confidenza, che potrebbero essere analizzati dall'utente mentre il sistema continua a raffinare l'analisi iniziale e quindi a fornire risultati a confidenza sempre crescente. In questo modo l'analista non deve attendere la fine della computazione, ma è coinvolto nel processo di analisi molto prima e con tempi di risposta compatibili con l'esplorazione e l'interazione umana. Lo sviluppo, la ricerca e l'applicazione alle reti sociali di soluzioni di analisi visuale progressiva (*Progressive Visual Analytics*<sup>4</sup>) possono sicuramente portare un contributo per le analisi più complesse e computazionalmente costose sulle reti sociali.

4. ANGELINI ET AL., 2018.

Infine, l'analisi visuale ha cominciato negli ultimi anni a essere utilizzata anche come supporto per analisi di tipo predittivo (ad esempio, di nuovi collegamenti o cammini di diffusione all'interno delle reti sociali) e di tipo speculativo (*what-if analysis*) nelle quali l'analista può ipotizzare lo stato di alcuni elementi della rete e ottenere una proiezione dello stato futuro rispetto alle ipotesi formate: un esempio di questa tecnica si può trovare in applicazioni di *disaster management* alimentate con dati provenienti dalle reti sociali.

In conclusione, i nuovi scenari di analisi delle reti sociali pongono sfide sempre più importanti e l'analisi visuale può rappresentare il giusto metodo per sfruttare sia la capacità computazionale dei moderni calcolatori che le tecniche di analisi automatica più moderne (intelligenza artificiale, apprendimento automatico) e le capacità di ragionamento umane al fine di implementare complessi pattern di analisi per l'estrazione di conoscenza e lo sfruttamento di questi dati che, per larghe porzioni, restano a oggi solo parzialmente utilizzati 

#### BIBLIOGRAFIA

- M. ANGELINI ET AL., *A Review and Characterization of Progressive Visual Analytics*, «Informatics» V (2018) 31, pp. 1-27.
- S. CHEN ET AL., *Social Media Visual Analytics*, «Computer Graphics Forum» XXXVI (2017) 3, pp. 563-587.
- K.A. COOK – J.J. THOMAS, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, «No. PNNL-SA-45230. Pacific Northwest National Lab. (PNNL)», Richland (2005).
- A. GOSWAMI – A. KUMAR, *Challenges in the Analysis of Online Social Networks: A Data Collection Tool Perspective*, «Wireless Personal Communications» XCVII (2017) 3, pp. 4015-4061.