

RETI TEMPORALI

Giorgio Locicero



Corso di Data Mining
a.a.2019-2020

INDICE

- Che cosa sono le reti temporali?
- Costruzione di una rete temporale
- Funzionamento tipico di una rete temporale
- Tipi di reti temporali
- Rappresentazione e teoria
 - 1)Raggiungibilità, insieme di influenza ed insieme dei nodi sorgente
 - 2)Distanza, durata, latenza, latenza di informazione e vector clock
 - 3)Grafo connesso transitivamente e latenza media
 - 4)Diametro, efficienza ed ambiguità
 - 5)Spanning tree e ambiguità sulle misure di centralità
- Possibili definizioni di adiacenza
- Misure di centralità e tendenza temporale
- Rappresentazioni di reti temporali come statiche
- Modellizzazione

Cosa sono le reti temporali?

- **Aggiunta di dettagli e attributi aggiuntivi che aumentano la possibilità di rappresentare reti reali, in particolare rappresentazione di reti dinamiche (neurali, internet, infezioni, informazioni e passaggio di informazioni tra nodi).**
- **Come per altri tipi di reti che approssimano situazioni reali(come multigrafi e grafi multidimensionali), nelle reti temporali si cerca di rappresentare la componente temporale della realtà e l'interazione intervallare tra varie entità.**
- **Molto importanti per rappresentare situazioni dove avviene un processo di espansione temporale e a stati.**
- **Cosa cambia rispetto alle reti statiche?**

Costruzione di una rete temporale

- **Una rete temporale non è creata direttamente dalla rete statica da analizzare, piuttosto si prende l'infrastruttura di base e viene fatta una creazione progressiva della rete temporale(dinamica), in questo modo possiamo attuare una analisi ibrida che tiene conto sia della rete statica che delle sue componenti dinamiche.**
- **Solitamente, viene messa in pratica una analisi unica della rete dinamica(rappresentata e trasposta opportunamente per le proprie esigenze), quindi durante la creazione si perdono alcune proprietà della rete statica.**
- **Bisogna trovare il giusto compromesso per non avere troppa complessità**

Costruzione di una rete temporale

- **La costruzione e l'analisi di una rete temporale sono pratiche che coinvolgono diverse discipline, una analisi topologica semplice potrebbe non essere abbastanza.**
- **Bisogna tenere conto di molte variabili, ed informazioni aggiuntive durante la costruzione.**

Funzionamento tipico di una rete temporale

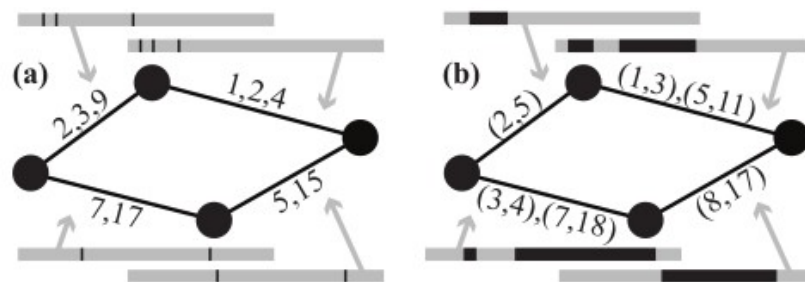
- **Ogni link (arco) di una rete temporale trasmette-trasporta informazioni solo durante il tempo in cui è attivo**
- **Questo tempo di attivazione può essere descritto da un quanto stabilito o da un intervallo di tempo adottando un approccio reale.**

Funzionamento tipico di una rete temporale

- **Si usano quanti di tempo ben definiti quando più che un approccio temporale puro, abbiamo un approccio a stati con passaggio di stati in tempi ben definiti (unici).**
- **Si usano intervalli di tempo in quasi tutte le situazioni che coinvolgono interazioni reali**
- **Solitamente si studiano reti reali basate su intervalli, ma le reti dinamiche basate su contatti sono più semplici da utilizzare ed analizzare**

Esempio di reti

- Nella rete basata su contatti (definiti da quanti di tempo) ogni arco possiede uno o più tempi di contatto
- Nella rete basata su intervalli ogni nodo possiede uno o più intervalli in cui è attivo



Possiamo vedere questo esempio come una possibile interazione tra diverse aree del cervello, nel primo caso scandiamo un periodo di sampling delle attivazioni, nel secondo caso prendiamo direttamente gli intervalli di attivazione.

Una ricerca di motifs può essere più semplice nel primo caso ricollegandoci a reti statiche.

Una analisi più reale può essere invece fatta dalla rappresentazione intervallare.

Tipi di reti temporali

- **Comunicazione e scambio di informazioni**
- **Interazioni uno a molti**
- **Metabolismo**
- **Sistemi distribuiti**
- **Reti neurali e cerebrali (di particolare importanza per analisi di microreti con contatti tra neuroni, o per analisi a grana grossa come l'attivazione di certe aree del cervello e la loro comunicazione, stesso discorso si può fare con le reti neurali)**
- **I possibili punti di applicazione sono infiniti.**

Rappresentazione e teoria

La rappresentazione della rete temporale parte dall'analisi della rete statica che prende come infrastruttura, abbiamo un grafo $G=(V,E)$ dove V è l'insieme dei nodi e E è l'insieme degli archi.

Come già visto in precedenza, possiamo rappresentare i grafi come contatti o come liste di intervalli associati ad ogni arco.

Nel caso dei contatti rappresentiamo ogni interazione come una tripla (i,j,t) dove i e j sono i nodi sorgente e destinazione e t è il tempo in cui avviene il contatto.

Questo tipo di approccio favorisce la visione della rete temporale come una serie di diapositive (snapshot) della rete in certi tempi, in modo da vedere la rete temporale come una sequenza di reti statiche

Rappresentazione e teoria

Altro tipo di rappresentazione che si avvicina di più a situazioni reali è quella ad intervalli di tempo.

Ad ogni nodo è associato un insieme di intervalli di tempo disgiunti ed ordinati in ordine crescente(per evitare ridondanze e ambiguità)
 $T = \{(t_1, t_1') , \dots (t_n , t_n')\}$

Per questo tipo di rappresentazione possiamo seguire l'approccio dei contatti e ricollegarci a reti statiche.

In questo caso però, più che definire una sequenza di reti statiche possiamo piuttosto definire una specie di matrice di adiacenza tridimensionale dove la terza dimensione è il tempo(se gli intervalli appartengono all'insieme $N \times N$ ovviamente)

$$a(i, j, t) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected at time } t \\ 0 & \text{otherwise} \end{cases}$$

Rappresentazione e teoria

- **Esistono anche altri tipi di rappresentazione(per esempio possiamo aggiungere il fatto per cui l'attraversamento di un arco comporta un tempo)**
- **Possiamo unire il concetto di rete temporale ad altri tipi di reti complesse ed ottenere qualcosa che si avvicina alla realtà ma cresce di complessità esponenzialmente**

Rappresentazione e teoria

Possiamo e dobbiamo definire i concetti e le metriche delle reti statiche e modificare le loro definizioni in modo da rispettare la componente temporale.

Definiamo il concetto di cammino tra due nodi

Un cammino è una sequenza ordinata di archi in cui i tempi (o gli intervalli) rispettano le regole stabilite sull'ordinamento temporale, in questo caso il cammino si dirà time-respecting-path(e.g. (A, B, 7), (B, C, 8), (C, D, 11)), da ora in poi con cammino intenderemo time respecting path per evitare ridondanza.

Importante notare che se esiste un cammino dal nodo i al nodo j e dal nodo j al nodo k, non è detto che esista il cammino tra il nodo i e k, quindi non vale la proprietà transitiva.

Raggiungibilità, insieme di influenza ed insieme dei nodi sorgente

Definiamo il concetto di raggiungibilità molto similmente a quello delle reti statiche:

Un nodo è raggiungibile da un altro nodo se e solo se esiste un time-respecting-path dal nodo sorgente a quello destinazione.

Definiamo l'insieme di influenza come l'insieme dei nodi raggiungibili dal nodo sorgente

Alternativamente possiamo definire l'insieme di tutti i nodi sorgente da cui è possibile raggiungere il nodo considerato.

Questi due insiemi possono essere visti come il passato ed il futuro del nodo

Raggiungibilità, insieme di influenza ed insieme dei nodi sorgente

Ovviamente anche la proprietà simmetrica non vale nelle reti temporali.

I concetti di connettività forte e debole vengono adattati alle reti temporali sempre tenendo conto della definizione di cammini e delle proprietà della rete insieme alla sua rappresentazione.

La ricerca di cliques è molto importante ma la maggior parte delle volte abbastanza complessa.

Distanza, durata, latenza, latenza di informazione e vector clock

La definizione di distanza resta la stessa delle reti statiche (sempre considerando la definizione di cammino)

Viene introdotta una nuova grandezza che è la durata, definita in un cammino come la differenza del tempo di arrivo dell'informazione e del tempo di inizio trasmissione dell'informazione

La latenza è la durata minima tra tutti i cammini tra due nodi

Definiamo la quantità $\phi_{i,t}(j)$ come l'ultimo tempo prima di t in cui l'informazione partita da j sia arrivata a i

Definiamo information latency come $\lambda_{i,t}(j) = t - \phi_{i,t}(j)$, misura di quanto è vecchia l'informazione.

Definiamo inoltre $[\phi_{i,t}(1), \dots, \phi_{i,t}(N)]$ come il vector clock

Grafo connesso transitivamente e latenza media

Definiamo sottografo connesso transitivamente se per ogni tripla $(V1, V2, V3)$ di nodi del sotto grafo se esistono dei cammini che congiungono $V1$ e $V2$, e $V2$ e $V3$, allora esisterà un cammino che congiunge $V1$ e $V3$.

Un uso molto comune dei concetti di durata e latenza in una rete temporale è il calcolo e l'approssimazione della “velocità” media di una rete, un carattere molto importante per lo studio ed il mining delle reti temporali.

Il problema del calcolo della latenza e durata media non è banale, la rete è dinamica e abbiamo più possibili cammini che rispettano l'ordinamento temporale da poter percorrere.

Diametro, efficienza ed ambiguità

La definizione di diametro è talvolta qualcosa che può essere descritto soltanto da decisioni personali e riguardanti le caratteristiche della rete temporale considerata (tempo massimo, considerazione metrica, caratteristiche interne, scatti o intervalli temporali, etc...)

Stessi problemi di definizione sorgono per la definizione di efficienza (da alcuni definita come media degli inversi delle distanze, da altri usando diverse metriche)

Una definizione accettata di diametro è quella del numero che anche se viene incrementato, non esistono cammini che colleghino due nodi del grafico che abbiano distanza minore, ma talvolta questa definizione viene invertita da alcuni autori.

Spanning tree e ambiguità sulle misure di centralità

Di grande importanza anche per le reti temporali è la costruzione del minimo spanning tree, sempre definito secondo le definizioni di raggiungibilità delle reti temporali.

Misure di centralità usate nelle reti statiche ricadono comunque negli stessi problemi riscontrati durante la definizione di diametro, principalmente per il fatto che molte misure usate nel graph mining sfruttano il concetto di adiacenza, quindi per riuscire a avere delle controparti nelle reti temporali dobbiamo definire il concetto di adiacenza in base a molte variabili

Possibili definizioni di adiacenza

Definizioni di adiacenza:

- normale
- Δt -adjacent, considerando solo archi che rientrino in un certo intervallo temporale

Tramite queste definizioni possiamo finalmente definire i temporal motifs come classi di isomorfismi.

Alcune interpretazioni deducono che la ricerca di isomorfismi sia riconducibile alla ricerca di isomorfismi in reti statiche, con conseguente scrematura con le adiacenze definite nella rete dinamica (vedere se effettivamente i nodi comunicano tra di loro nella rete temporale), con conseguenti modi di fare questa scrematura (diversi tempi, prendere tutte le possibilità o solo quelle più moderne)

Misure di centralità

La closeness centrality è definita per le reti statiche come

$$C_C(i) = \frac{N - 1}{\sum_{j \neq i} d(i, j)},$$

Dove $d(i, j)$ è la distanza geodesica tra due nodi del grafo, ed N è il numero di nodi del grafo

Se il nostro studio della rete dinamica è incentrato sullo studio della velocità con cui l'informazione arriva ai nodi possiamo definire la nostra closeness centrality come

$$C_C(i, t) = \frac{N - 1}{\sum_{j \neq i} \lambda_{i, t}(j)},$$

Dove $\lambda_{i, t}(j)$ è l'information latency descritta già in passato

Misure di centralità

Possiamo definire un'altra misura di closeness che tenga conto dell'efficienza della rete e degli intervalli infiniti (nodi irraggiungibili entro il tempo t):

$$C_E(i, t) = \frac{1}{N-1} \sum_{j \neq i} \frac{1}{\lambda_{i,t}(j)}$$

Dove $1/\lambda_{i,t}(j)$ è definito come 0 se non ci sono cammini da j a i entro il tempo t

Un'altra misura importante basata sui cammini minimi è la betweenness centrality definita per reti statiche ma facilmente trasponibile a reti dinamiche, che misura il numero di cammini minimi che passano per il nodo considerato

$$C_B(i) = \frac{\sum_{j \neq k} v_i(j, k)}{\sum_{j \neq k} v(j, k)}$$

Dove $v_i(j, k)$ è il numero di cammini minimi che passano per i, e $v(j, k)$ è il numero totale di cammini minimi tra j e k

Misure di centralità e dove approfondire

Esistono altre misure di centralità ed altri modi di interpretare le varie metriche che vengono usate per calcolarle, per chi vuole approfondire ancora di più l'argomento leggere

Path lengths, correlations, and centrality in temporal networks. R. K. Pan and J. Saramäki.

Ed altri documenti riguardanti reti reali che vengono studiate e analizzate(quasi tutti i documenti a fine [1] usano misure abbastanza comuni o ridefinite in base a come venivano definite ed interpretate le componenti temporali)

Altre misure importanti

La burstiness che esprime il comportamento effettivo di molte reti reali, che alternano trasmissioni di informazioni solo in brevi tratti di tempo in cui la rete si intasa di informazioni, con tratti di completa inattività nella rete.

Altre misure molto usate sono quelle che usano il concetto di informazione, che è alla base di molte reti temporali, quali entropia ed altri metodi.

Rappresentazioni di reti temporali come statiche

La letteratura sulle reti statiche (grafi o varianti statiche) risulta essere immensamente più grande e dettagliata, principalmente perché sono molto più semplici da analizzare.

Infatti uno degli approcci nella rappresentazione e nell'analisi di reti temporali è quello di usare reti che rispecchino sia le proprietà delle reti statiche che proprietà dinamiche (proprietà temporali e topologiche).

Per esempio il Reachability Graph è molto importante nella rappresentazione di grafi per connessioni sparse. Si costruisce un arco direzionato tra due nodi se esiste un cammino tra i due nodi.

Si può dimostrare che per ogni rete temporale si possono costruire uno o più reachability graph, e si può dimostrare che per ogni reachability graph si può costruire una rete temporale che lo rispetti.

Rappresentazioni di reti temporali come statiche

Altro modo di rappresentare le reti dinamiche come reti statiche usa il concetto di line graph, dove il ruolo dei nodi e degli archi viene invertito.

Viene costruito quindi un transmission graph, che incorpora nella definizione dei nodi e degli archi del line graph anche la componente temporale(sia del momento del contatto, sia per quanto tempo avviene)

Modellizzazione di reti temporali

Il concetto di modellizzazione di reti temporali assume significati diversi da quelli della statistica inferenziale, dove quando modellizziamo tentiamo di costruire un modello per predire e studiare valori e parametri.

La modellizzazione di reti temporali si avvicina di più alla creazione di reti che abbiano certi comportamenti o che siano utili per simulazioni e studi di comportamenti particolari simulando reti reali

Modellizzazione di reti temporali

Durante l'analisi , confrontiamo le reti temporali reali con modelli randomizzati creati ad hoc per carpire feature che a primo sguardo non possiamo vedere.

Analizziamo parti della rete in certi intervalli temporali.

Creazione di reti temporali

La creazione della rete temporale randomica parte da una ipotesi iniziale (da questo il modello prende il nome di null model dato che dell'ipotesi costruiamo attorno le nostre congetture, ma stabiliamo comunque un certo grado di randomicità), che può essere per esempio il grado medio di ogni nodo, la distribuzione di intervalli medi di attivazione degli archi, o la probabilità di effettuare contatti tra nodi in certi intervalli temporali.

Linee guida durante la creazione di reti temporali

Non esiste un modello generalizzato di reti temporali randomizzate, data la natura arbitraria dei null models, piuttosto si possono o seguire delle linee guida sapendo la struttura base della rete che vogliamo analizzare, oppure possiamo andare di forza bruta fino ad esaurimento delle possibilità, cioè provare diversi null models e vedere quale tra questi ci dà i risultati più soddisfacenti.

Seguendo una struttura di base, possiamo prenderlo ad esempio come infrastruttura e seguire strutture topologiche e temporali.

Modelli randomici per reti temporali

Solitamente I modelli randomici usati sono modelli generativi che partono da una rete temporale di partenza e modificano la struttura temporale e/o topologica per raggiungere un certo livello sia di randomicità che di regolarità(attributi che vanno in trade-off)

Modelli randomici per reti temporali

Randomized edges (RE), dove cambiamo le destinazioni con probabilità p

Randomly permuted times per la controparte temporale, la struttura statica viene mantenuta, i tempi di contatto rimangono lo stesso numero ma vengono permutati tra i vari archi, la cardinalità dei tempi di contatto rimane uguale

RE + RP è un ibrido tra i due modelli.

RP conserva l'insieme dei tempi di contatto, ma distrugge il comportamento tipico delle reti temporali di burstiness, un altro modello che distrugge ancora di più questi pattern di comportamento è il **random times (RT)** dove vengono scelti tempi randomici per ogni contatto, ovviamente la generazione dei tempi random non deve seguire per forza una distribuzione uniforme.

Modelli randomici per reti temporali

Random contacts(RC) per mantenere la topologia e ridistribuire i contatti.

Equal weight edge randomization(EWER) archi con lo stesso numero di contatti possono essere sostituiti. Questo modello mantiene la burstiness, ma ha bisogno di una rete abbastanza grande da avere abbastanza contatti.

Edge randomization(ER) è la generalizzazione in cui possiamo scambiare qualsiasi arco. La topologia della rete distrutta, la temporalità relativamente conservata.

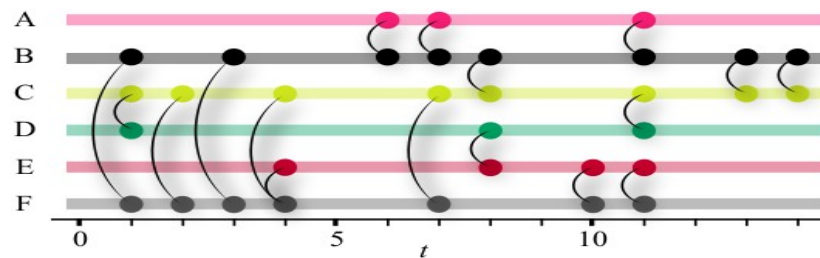
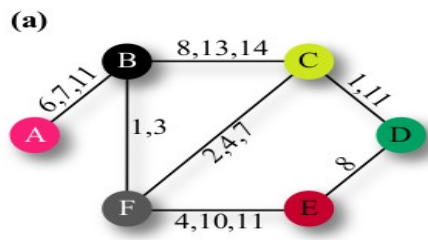
Time reversal(TR) con tempi inversi, cioè percorriamo i contatti in senso inverso, si cercano correlazioni tra la rete normale e quella inversa.

Decisione dei modelli randomici da utilizzare

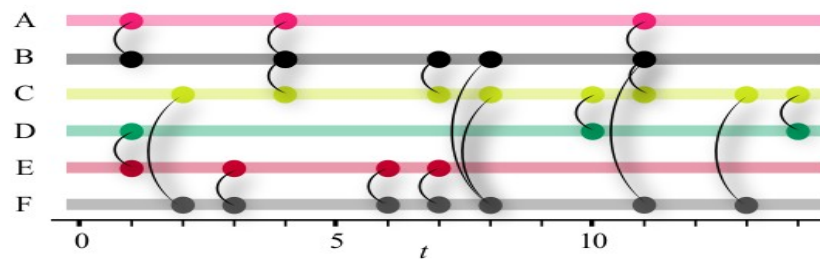
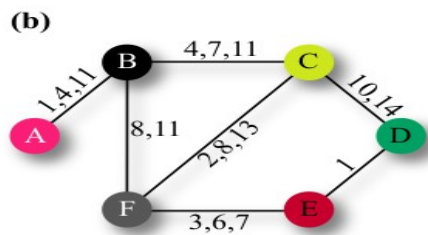
Euristico ed empirico è principalmente quasi tutto il procedimento di ricerca e di costruzione dei null models, ogni modello favorisce qualche caratteristica e distrugge le altre, solitamente si fa un'analisi sfruttando un gruppo di questi modelli in modo da analizzare una buona parte delle caratteristiche della rete temporale, oppure si sfruttano ibridi (e.g. RE + RP) che permettano di mantenere un equilibrio tra prevedibilità e casualità.

Per grafi con molti contatti il risultato di usare EWER è abbastanza soddisfacente, alternato con altri modelli risulta come uno dei “migliori” metodi adottati nell'analisi dei grafi

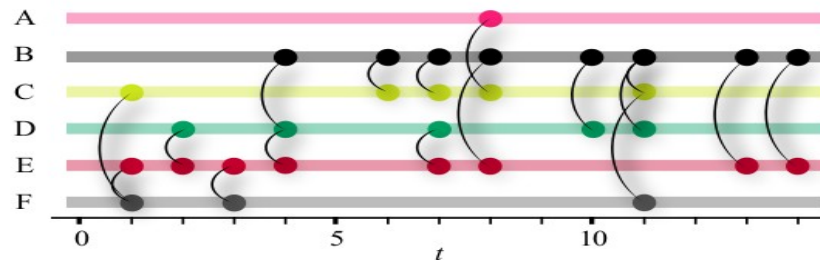
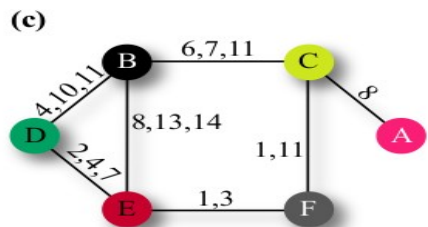
Esempi di modelli randomici per reti temporali



a) rete temporale di base rappresentata come sequenza di contatti



b) RP



c) RE

Dinamica dei processi di espansione e propagazione

Importante e di fondamentale comprensione è la parte della teoria delle reti temporali che studia i processi e le dinamiche di espansione e propagazione(spreading) di fenomeni, modelli compartimentale come suddivisione in stati caratteristici, rappresentati dai nodi della rete temporale, cioè ogni nodo appartiene ad un compartimento

Per esempio nel modello di espansione SIR un nodo può appartenere a tre compartimenti differenti, suscettibile, infetto e recovered(guarito), I passaggi da un compartimento ad un altro avvengono tramite funzioni che tengono conto del tempo di contatto tra I vari compartimenti)

Modello di espansione SIR

La probabilità di infezione e di guarigione (infection rate e recovery rate) sono importanti parametri per il modello.

Applicare questo modello significa supporre che tutte le interazioni avvengano uniformemente nel tempo, cosa che non è vera la maggior parte delle volte.

Per esempio l'infezione di un virus avviene anche in base al tempo di contatto con l'infetto, stesso discorso può essere fatto sul passaggio da infetto a guarito.

Come controlliamo questo modello?

Vaccinazioni diminuiscono la probabilità di essere infettati, possiamo introdurre modelli di strategie di vaccinazione e di prevenzione.

Questo modello, come tutti i modelli che si rifanno a situazioni reali, assume durante le varie interazioni e passaggi di compartimento un comportamento bursty.

Problemi delle reti temporali

- **Prima e durante la costruzione di reti temporali, bisogna tenere conto di quante informazioni perdiamo e quanto valga la pena costruire una rete temporale, cioè considerare la fattibilità, l'accuratezza e la manutenzione della rete.**
- **Problemi di scalabilità, confronto della rapidità di cambiamento della rete statica e temporale.**
- **Visualizzazione e rappresentazioni di una rete potrebbero comprometterne sia la complessità che l'interpretazione**

Problemi delle reti temporali

- **Intransitività non permette di usufruire di tecniche basate su di essa (troppe)**
- **La letteratura scientifica sulle reti temporali è poca rispetto a quella sulle reti statiche**
- **La letteratura sulle reti temporali è molto sparsa, vi sono molte ambiguità durante lo studio dell'argomento, e solitamente si lascia la libertà di scelta ed interpretazione.**
- **Le reti temporali sono sistemi complessi particolarmente delicati e ingannevoli.**
- **Interpretazioni varie sia della natura temporale che dell'origine delle interazioni**

Problemi delle reti temporali

- Per essere significativa la nostra costruzione ed analisi bisogna avere delle interazioni binarie, cioè tra due nodi, in modo da mantenere sia un certo livello di prevedibilità, sia una certa randomicità nel comportamento della rete.
- Le reti temporali costruite non dovrebbero essere né troppo randomiche né troppo regolari per non allontanarsi troppo dalla rete statica a cui sono state interfacciate.
- Attenzione alla velocità dei contatti rispetto alla velocità di cambiamento del sistema dinamico.
- Essendo un sistema dinamico, la rete temporale è particolarmente difficile da analizzare, nodi possono rientrare ed uscire, archi possono non essere percorsi, bisogna tenere conto di molte variabili ed alcune potrebbero non essere intuibili.

Possibilità future e contenuti aggiuntivi

- **Nella ricerca di motifs, cliques, community, e altri tipi di pattern che hanno una certa regolarità esistono diverse interpretazioni.**
- **Importanti nell'analisi di reti temporali non sono soltanto le cause topologiche o pratiche per cui avvengano certe interazioni, anche lo studio delle origini delle interazioni assume rilievo, per esempio lo studio del contatto tra due nodi che, come già detto, assume un comportamento bursty più che seguire una distribuzione conosciuta(e.g. poisson)**

Possibilità future e contenuti aggiuntivi

- **Lo studio delle reti temporali potrebbe giovare di molti studi che vengono fatti sulle reti statiche(come per esempio la colorazione dei grafi, o la grande quantità di metodi utili per la ricerca dei cammini minimi, per la ricerca delle chiusure transitive)**
- **Generazione di reti con caratteristiche particolari**
- **Descrittori di struttura temporale che rispecchino la vera natura della componente temporale, che siano create e studiate per essere utili nell'analisi delle reti dinamiche**

Possibilità future e contenuti aggiuntivi

- **analisi delle community**
- **Ricerca di cluster**
- **Visione mesoscopica sia della struttura statica che dinamica**

FONTI

- 1) SURVEY ON TEMPORAL NETWORK di Peter Holme e Jari Saramäki (con le fonti citate a fine survey).
- 2) Temporal network in R (programminghistorian.org/en/lessons/temporal-network-analysis-with-r)
- 3) Mapping temporal-network percolation to weighted, static event graphs (www.nature.com/articles/s41598-018-29577-2)
- 4) Wikipedia