



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO



DIPARTIMENTO DI
INFORMATICA

**Dottorato di ricerca in Informatica e Matematica
XXXII ciclo**

Progetto di ricerca

Dottorando: Dott. Angelo Impedovo

Tutor: Prof. Michelangelo Ceci

Firma del dottorando _____

Firma del tutor _____

1. Titolo della ricerca:

Mining temporal patterns in large heterogeneous networks

2. Area nella quale si inquadra la ricerca:

Data Mining, Complex Networks Analysis, Big Data

3. Motivazioni e obiettivi della ricerca

Lo studio di sistemi complessi è attività necessaria in diversi campi scientifici e sociali. Da un punto di vista concettuale questi sistemi possono essere modellati per mezzo di complesse reti eterogenee, ossia reti con diversi tipi sia di nodi che di archi fra i nodi. L'indagine dei comportamenti esibiti da un sistema complesso, pertanto, può tradursi nell'analisi delle proprietà statistiche e strutturali della rete che lo modella.

In natura e in società esistono reti complesse che descrivono diversi fenomeni [1, 2] (per esempio reti sociali, reti di contatti, reti di contagio, reti di interazione metaboliche, ecc.), di contro una stessa rete può descrivere diversi fenomeni; alcune reti sono dinamiche e altre statiche a seconda del caso in cui la loro struttura possa variare o meno nel tempo. Lo studio delle reti complesse, originariamente inquadrato nell'ambito della teoria dei grafi [3], è oggi un settore scientifico in crescente espansione anche grazie al potere computazionale abilitato dalle tecnologie di High Performance Computing e Big Data.

Alcuni aspetti che è utile indagare possono riguardare lo studio della tolleranza ai guasti in una rete, lo studio della diffusione delle informazioni al suo interno, lo studio delle dinamiche di ramificazione, di frammentazione o deframmentazione e, in generale, lo studio della sua evoluzione complessiva.

In linea di massima si assume che il processo evolutivo di una rete dinamica sia guidato da specifici processi perlopiù sconosciuti agli osservatori, in questa accezione si parla di processi stocastici capaci di generare diverse immagini "istantanee" di una specifica rete in diversi punti temporali.

I principali obiettivi da conseguire nell'analisi delle reti dinamiche sono:

- Indurre modelli descrittivi accurati del processo evolutivo in base a quanto osservato.
- Indurre modelli predittivi capaci di anticipare le dinamiche del processo evolutivo.

Nel perseguire queste finalità, spesso, ci si scontra con le problematiche ben note in area Big Data: le immagini di rete possono essere prodotte con elevata velocità, la loro struttura può essere enormemente estesa e può arrivare a contemplare informazioni complesse, occorre inoltre verificare l'affidabilità di tali informazioni.

Scopo del progetto di ricerca sarà quello di affrontare l'analisi temporale delle reti in modo da sintetizzare modelli dei processi che ne regolano l'evoluzione, per far ciò occorrerà mettere a punto metodi automatici (afferenti al campo del data mining e dell'apprendimento automatico) capaci di produrre risultati statisticamente affidabili in una maniera efficiente e computazionalmente compatibile con gli scenari tipici del contesto big data.

Indagare entrambe le dimensioni di efficienza e robustezza dei modelli è cruciale. La questione dell'efficienza deve essere affrontata in virtù delle necessità di distribuire il carico computazionale,

di algoritmi appositamente progettati per il funzionamento su singola unità computazionale, su più unità di calcolo. Infine la robustezza statistica dei modelli prodotti deve necessariamente essere contemplata, pena l'invalidazione delle conclusioni ottenute.

In particolare ci si concentrerà sulla scoperta di modelli descrittivi capaci di rappresentare, entro una certa approssimazione, gli aspetti frequentemente occorrenti, i cambiamenti graduali e/o immediati, gli aspetti periodici e le eventuali correlazioni.

4. Stato dell'arte

Le reti dinamiche emergono in numerosi domini applicativi spesso molto differenti, si pensi alla rete dei documenti nel world wide web, alle reti sociali e alla rete delle comunicazioni. In maniera del tutto alternativa, si fa riferimento alle reti usando anche il termine "grafo"; soltanto recentemente, l'interesse nell'area delle reti sociali dinamiche ha portato un crescente interesse nello studio delle reti dinamiche di stampo generale [4].

Lo studio della evoluzione delle reti dinamiche comprende due fondamentali classi di metodi: i metodi manutentivi e quelli puramente analitici. Nei metodi manutentivi è desiderabile che i risultati del processo di data mining siano continuamente adattati nel tempo, ad esempio un modello di classificazione o di clustering deve anch'esso evolvere di pari passo con la struttura della rete osservata, un requisito fondamentale per questi metodi è quello di incrementalità. I metodi analitici, di contro, quantificano direttamente l'intensità del cambiamento osservato nella rete per mezzo di alcuni indicatori matematici.

Aspetto di interesse rilevante è quello per cui non tutte le reti evolvono con la stessa rapidità, alcune cambiano ad un livello di granularità temporale più fine (ad esempio ogni secondo, come nel caso della rete delle mail scambiate fra individui) mentre altre lo fanno ad un livello maggiore (ad esempio ogni settimana, come nel caso delle collaborazioni scientifiche fra autori). Differenti scenari a differenti livelli di granularità temporale possono richiedere differenti tipologie di analisi, fondamentalmente si individuano due famiglie di approcci: quelli per reti che evolvono lentamente e quelli per reti che evolvono in streaming.

Negli approcci per reti che evolvono lentamente si esaminano immagini istantanee complete della stessa rete osservate in punti temporali differenti, in questo contesto si parla anche di "snapshot analysis" e si preferisce espletare direttamente una analisi "offline" che non tenga conto di nuovi dati in arrivo. Gli approcci di streaming, contrariamente al primo caso, gestiscono reti costruite da una serie di interazioni transienti indipendenti (si pensi alle migliaia di mail fra loro indipendenti che, collettivamente, strutturano la rete di consegna della posta elettronica) che vanno analizzate in modalità "online". Scenari di streaming possono emergere nei contesti di stream di oggetti [5] e di archi [6] e sono di gran lunga più avvincenti vista l'impossibilità di contenere un intero grafo in memoria.

Come è facile evincere, la categorizzazione dei metodi e quella della velocità evolutiva delle reti danno luogo ad una serie di possibili combinazioni. Per esempio, sia i metodi analitici che quelli manutentivi possono essere studiati sia nel caso snapshot che nel caso streaming; nella parte che segue si cercherà di espandere il panorama di alcuni lavori ritenuti noti per quanto concerne lo studio del comportamento evolutivo di una rete.

Per quanto concerne i metodi manutentivi occorre distinguere fra approcci di tipo snapshot e streaming. Gli approcci di tipo snapshot includono metodi di clustering evolutivo, approssimazioni low rank e approcci di link prediction; si ritengono particolarmente importanti le tecniche di clustering poiché al loro interno ricadono approcci di pattern mining: Ahmed e Karypis [7] hanno

proposto un metodo che si occupa di utilizzare l'analisi di pattern relazionali capace di descrivere transizioni di stato in una rete, alternativamente Berlingerio et al. [8] hanno proposto un metodo in grado di scoprire regole, estendendo la nozione di pattern frequente, descrittive del comportamento evolutivo.

Negli approcci di tipo streaming, in cui un insieme potenzialmente illimitato di archi è continuamente in arrivo, si collocano tecniche capaci di supportare ingenti quantitativi di dati, in questi ricadono metodi di clustering e di dense pattern mining: Aggarwal et al. [5] hanno proposto un metodo capace di raggruppare piccoli grafi utilizzando una speciale tecnica di partizionamento, agendo a questa maniera è possibile rappresentare efficientemente il grafo anche se in forma approssimata; In [9] si è affrontato il problema di fornire una implementazione distribuita capace di estendere l'attività di clustering.

Gli approcci di dense pattern mining [5] diventano quindi importanti perché, a conti fatti, un cluster non è altro che un pattern molto denso proveniente dai dati: un denso gruppo di nodi è definibile come un gruppo di nodi che co-occorrono frequentemente tale da avere elevata densità di archi.

Tuttavia, i metodi snapshot non sono facilmente adattabili al contesto dei dati forniti in streaming. Per tale ragione lo studio di metodi in grado di estrarre pattern temporali a partire da grandi reti eterogenee è considerata un'area di ricerca molto fertile. Spesso si cerca di convertire un approccio snapshot in uno streaming utilizzando tecniche basate su finestre temporali, come approfonditamente studiato da Gama [10]; i modelli a finestre temporali, solitamente di tipo sliding (finestre di dimensione fissa che traslano lungo l'asse temporale), landmark (finestre di dimensione crescente) e tilted (finestre di dimensione crescente in cui l'importanza degli oggetti è pesata rispetto all'istante di provenienza) offrono un valido aiuto per la creazione di metodi incrementali e sufficientemente efficienti.

Per quanto concerne i metodi analitici occorre distinguere anche qui fra approcci di tipo snapshot e streaming, occorre tuttavia fare una doverosa premessa: in queste tecniche lo scopo principale è quello di comprendere le dinamiche globali dell'evoluzione, non si cerca di mantenere continuamente aggiornato un modello. La conseguenza della distinzione è l'esistenza di un profondo legame fra gli approcci analitici e quelli manutentivi, infatti spesso è necessario basare alcune decisioni dei secondi sulle conclusioni tratte dai primi.

Un esempio di metodo analitico in un contesto snapshot è quello di descrivere l'evoluzione della rete mediante uno studio delle distanze minime fra i nodi, Ley [11] ha dimostrato che l'apparizione di un arco fra due nodi enormemente distanti è un evento inusuale nella rete delle collaborazioni bibliografiche (DBLP), eventi di questo tipo è probabile che riflettano collaborazioni fra autori di settori differenti.

Per quanto concerne uno scenario tipicamente streaming è possibile far riferimento alla attività di anomaly detection presentato da Aggarwal et al. [12], nel lavoro si decide di accettare o rigettare un nuovo arco in base alla probabilità stimata (a priori) di vederlo apparire.

La possibilità di rappresentare una rete in maniera relazionale rende ulteriormente interessanti le attività di pattern mining, è possibile infatti ricorrere a quanto abbondantemente studiato in letteratura sulla scoperta dei pattern frequenti [13] e sulle attività di graph mining [14]. I modelli descrittivi intesi come insiemi di pattern possono anche essere utilizzati per costruire solidi modelli predittivi, un esempio è il lavoro proposto da Dong et al. [15] in cui si costruisce un classificatore associativo sulla base di pattern emergenti [16] estratti a partire dai dati. In analogia con quanto visto in precedenza, anche la scoperta dei pattern è un task manutentivo: è necessario adattare l'insieme dei pattern ritrovati (siano essi semplicemente frequenti o di altro tipo) in modo da

rispondere a quanto osservato in un contesto sia snapshot che streaming. Per quanto riguarda le attività di pattern mining applicate a grafi, Calders et al. [17] hanno proposto funzioni di valutazione del supporto di un pattern specificatamente progettate per algoritmi di graph mining.

5. Approccio al problema

Nel progetto si intende affrontare la problematica di estrarre pattern temporali a partire da grandi reti eterogenee in modo da catturarne l'evoluzione; in reti di questo tipo sia i nodi che gli archi possono essere di tipo differente, inoltre sia i nodi che gli archi possono apparire o sparire lungo l'asse temporale: le tecniche che si intende mettere a punto non possono quindi prescindere dalle problematiche sia di aggiornamento e incrementalità del modello (tipiche delle tecniche manutentive viste in precedenza) che da quelle di gestione di reti molto ampie (tipiche del contesto in streaming). Si intende fornire soluzioni computazionalmente efficienti, in linea con quanto studiato in [10].

Nell'accezione data, un pattern temporale è una generica feature multivariata osservata lungo l'evoluzione della rete che veicola uno specifico significato. Potrebbe essere molto interessante scoprire aspetti ben consolidati per mezzo dei pattern frequenti, cercare di scoprire cambiamenti per mezzo di pattern emergenti o, piuttosto, aspetti periodicamente ricorrenti.

La modellazione dell'evoluzione per mezzo dei pattern abilita a diversi vantaggi: in primo luogo ogni pattern non sarebbe altro che una sottorete di quella osservata e finirebbe per dividerne lo stesso linguaggio di rappresentazione, in secondo luogo così facendo sarebbe possibile integrare il processo di ricerca con gli aspetti analitici studiati per l'analisi delle reti complesse, in aggiunta a questo sarebbe sempre possibile decidere di mantenere costantemente aggiornato il modello nel tempo. L'attività di imporre vincoli aggiuntivi all'attività di scoperta dei pattern prende il nome di constrained frequent pattern mining [18]. Un esempio banale potrebbe essere quello di scoprire pattern periodici imponendo il vincolo aggiuntivo di elevata densità o diametro, diversamente da quanto studiato in [5, 8, 13].

Prospettiva interessante è offerta dalla possibilità di impiegare un modello descrittivo, espresso come insieme di pattern, per migliorare le performance di un modello predittivo, tipicamente un classificatore. La ricerca, rispetto a quanto fatto in [15], abiliterebbe all'uso di strategie più sofisticate per la costruzione del modello predittivo.

Problematica connessa alla scoperta di pattern è quella di gestirne l'elevata numerosità. Spesso il numero di pattern estratti è talmente grande da mettere in difficoltà qualunque utente del task di data mining. Una prospettiva che si intende esplorare è quella di organizzare tali insiemi di pattern temporali in modo da poter costruire dei modelli computazionali generativi (ad esempio in forma di grammatiche a grafo) capaci di simulare il processo stocastico che guida l'evoluzione.

6. Risultati attesi

Le applicazioni possono interessare diversi domini applicativi, si pensi allo studio delle reti sociali, delle reti stradali, di quelle dei contagi, delle reti di comunicazione e via discorrendo; nonostante un interesse diffuso e riconosciuto per quanto riguarda lo studio dell'evoluzione delle reti, le ricadute applicative dei risultati attesi dal progetto mirano ad essere competitive vista e considerata la loro pertinenza con il contesto delle analisi in streaming e, per estensione, per la loro richiesta di un apparato tecnologico compatibile con lo scenario dei Big Data. Gli algoritmi e le metodologie che si intende sviluppare potrebbero includere aspetti tipici del calcolo distribuito o, più in generale, del

calcolo ad elevate prestazioni.

Numerosi sono i casi applicativi, ad esempio l'analisi dell'evoluzione delle rete impatta fortemente sullo studio delle reti di comunicazione, nello specifico Chan et al. [19] hanno affrontato il problema di individuare cambiamenti correlati spazio-temporalmente in grandi reti. Lo studio è necessario in modo da poter analizzare la diffusione dei guasti all'interno della rete e poterne identificare le cause reali. Il lavoro di Mongiovi et al. [20] si occupa invece di indagare l'evoluzione del traffico stradale; in particolare si modella la rete stradale come una rete pesata, l'intensità del traffico è strutturata come pesi numerici degli archi: in questo senso è possibile studiare l'evoluzione dei pesi nel tempo in modo da poter identificare variazioni significative nei flussi di traffico. Nei lavori di Vasques et al. [21] e di Solé et al. [22] si modellano delle reti di interazione fra proteine, in cui i nodi sono le proteine e gli archi corrispondono a interazioni fra loro, così facendo si studia l'evoluzione lungo l'arco di vita delle proteine; un altro caso interessante di applicazione relativa allo studio delle reti biologiche è quello offerto da Beyer et al [23], in diverse patologie, come quella del diabete di tipo 2, rotture nei percorsi che coinvolgono relazioni con l'insulina influenzano enormemente l'evoluzione della rete; la scoperta di specifici pattern di evoluzione può portare a differenti diagnosi. Un forte interesse industriale, oltre che dal campo biologico, può provenire da quello dei dispositivi mobile e delle telecomunicazioni; nel lavoro di Akoglu et al. [24] si osserva come delle proprietà strutturali importanti della rete riescano ad influenzare la persistenza di link fra individui, spesso l'evoluzione delle interazioni nelle reti di dispositivi mobile corrisponde a eventi capaci di modificare le interazioni fra gli individui (festival, riunioni, ecc.).

In definitiva è possibile affermare che la scoperta di pattern temporali delle reti dinamiche, per spiegare la loro evoluzione, può trovare numerose applicazioni sia a livello scientifico che industriale. Inoltre (tendenzialmente) il framework della teoria delle reti complesse fornisce un livello di astrazione cui rivolgere l'attenzione e può essere proficuamente utilizzato per lo studio del comportamento di ogni sistema complesso rappresentabile con reti dinamiche eterogenee.

7. Fasi del progetto

Il progetto è da intendersi articolato in 3 anni di tempo con diverse fasi per anno:

- Primo anno: studio preliminare della letteratura e dello stato dell'arte
 - Attività 1A: studio della letteratura e dello stato dell'arte.
 - Attività 1B: approfondimento delle tematiche relative alle tecniche di estrazione di pattern temporali
 - Attività 1C: ricerca e studio di metodi per l'analisi delle reti dinamiche.
 - Attività 1D: indagine circa l'attività svolta da gruppi di ricerca con obiettivi affini.
- Secondo anno: sintesi, realizzazione e implementazione dei metodi
 - Attività 2A: partecipazione a scuole internazionali inerenti al tema di ricerca.
 - Attività 2B: sintesi, progettazione e implementazione di metodi per la scoperta di pattern temporali da grandi reti eterogenee, in linea con gli obiettivi previsti.
 - Attività 2C: valutazione dei metodi realizzati, confronto con approcci esistenti e pubblicazione dei risultati conseguiti in riviste e conferenze internazionali.
- Terzo anno: applicazioni pratiche e redazione della tesi di dottorato
 - Attività 3A: periodo formativo presso università straniera e confronto con l'attività svolta presso altri gruppi di ricerca con obiettivi affini.
 - Attività 3B: raffinamento dei metodi e realizzazione di caratteristiche specifiche per i domini applicativi scelti.
 - Attività 3C: analisi dei risultati sperimentali ottenuti sui domini applicativi selezionati.

- Attività 3D: redazione della tesi finale di dottorato.

8. Valutazione dei risultati

La valutazione dei risultati avverrà usando specifiche metriche a seconda di quelli che saranno gli scopi dei metodi proposti, in seguito se ne indicheranno sommariamente alcune. Nel caso in cui si voglia esclusivamente valutare la bontà del modello descrittivo occorre tenere conto di alcune semplici statistiche descrittive sui pattern, le più conosciute sono il supporto relativo e la copertura. Considerando un dataset D , la cui generica istanza è identificata come $i \in D$, allora un generico pattern P è assimilabile a una funzione binaria $P: D \rightarrow \{\text{vero}, \text{falso}\}$. Un oggetto $i \in D$ si dice coperto da P se $P(i) = \text{vero}$. La copertura di P è pari all'insieme degli oggetti coperti, il supporto relativo di P è pari al rapporto fra la cardinalità della sua copertura e la cardinalità del dataset.

$$\text{Copertura}(P) = \{i \in D \mid P(i) = \text{vero}\} \quad \text{Supporto}(P) = \frac{|\text{Copertura}(P)|}{|D|}$$

Nel valutare complessivamente le proprietà di un modello si terrà conto di alcune semplici statistiche descrittive come il numero di pattern estratti, il supporto medio, il campo di variazione del supporto e la deviazione standard del supporto. Lo studio delle reti complesse è in grado di fornire un framework per la valutazione strutturale dei pattern, infatti visto che ogni pattern è espresso come un grafo $P = (V, E)$, ove V è l'insieme dei vertici e $E \subseteq V \times V$ è l'insieme degli archi, allora diventa possibile quantificare alcune statistiche strutturali:

$$\text{Ordine}(P) = |V| \quad \text{Dim}(P) = |E|$$

$$\text{Grado}(P) = 2 \text{Dim}(P) \quad \text{Densità}(P) = \frac{\text{Grado}(P)}{\text{Ordine}(P)(\text{Ordine}(P)-1)}$$

Nel valutare complessivamente le proprietà strutturali di un modello graph-based è possibile fornire una declinazione statistica degli indicatori di cui sopra (grado medio, densità media, ordine medio, dimensione media, ecc.). Nel decidere se la bontà di un modello descrittivo deponga a favore delle performance di un modello predittivo si farà ricorso ad alcune metriche di valutazione della accuratezza predittiva:

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN} \quad \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad F_1 = \frac{2TP}{2TP + FP + FN}$$

La valutazione permette di osservare un numero di veri positivi TP e veri negativi TN (ossia istanze correttamente classificate dal modello, rispettivamente di classe positiva o negativa) ma anche dei falsi positivi FP e falsi negativi FN (ossia istanze non correttamente classificate dal modello, rispettivamente di classe positiva o negativa). Valutazioni di precision e recall quantificano la bontà di un modello predittivo rispetto ai falsi positivi e ai falsi negativi osservati, metriche più complesse (come nel caso di accuracy e F1-score) cercano di mediare tali risultati.

9. Eventuali referenti esterni al Dipartimento

Per quanto concerne il coinvolgimento di eventuali referenti esterni al dipartimento si segnalano:

- Toon Calders, Université Libre de Bruxelles (ULB), Computer & Decision Engineering (CoDE) department, Web and Information Technologies (WIT) group.
- Joao Gama, University of Porto, Faculty of Economics, Laboratory of Artificial Intelligence and Decision Support (LIAAD – INESC TEC)

10. Riferimenti bibliografici

1. Strogatz, Steven H. "Exploring complex networks." *Nature* 410.6825 (2001): 268-276.
2. Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.
3. Erdos, Paul, and Alfréd Rényi. "On the evolution of random graphs." *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960): 17-60.
4. Aggarwal, Charu C. "An introduction to social network data analytics." *Social network data analytics*. Springer US, 2011. 1-15.
5. Aggarwal, Charu C., et al. "On dense pattern mining in graph streams." *Proceedings of the VLDB Endowment* 3.1-2 (2010): 975-984.
6. Zhao, Peixiang, Charu C. Aggarwal, and Min Wang. "gSketch: on query estimation in graph streams." *Proceedings of the VLDB Endowment* 5.3 (2011): 193-204.
7. Ahmed, Rezwana, and George Karypis. "Algorithms for mining the evolution of conserved relational states in dynamic networks." *Knowledge and Information Systems* 33.3 (2012): 603-630.
8. Berlingerio, Michele, et al. "Mining graph evolution rules." *joint European conference on machine learning and knowledge discovery in databases*. Springer Berlin Heidelberg, 2009. Aggarwal, Charu C., Yuchen Zhao, and S. Yu Philip. "Outlier detection in graph streams." *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*. IEEE, 2011.
9. Stanton, Isabelle, and Gabriel Kliot. "Streaming graph partitioning for large distributed graphs." *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012.
10. Gama, Joao, and Mohamed Medhat Gaber. *Learning from data streams*. Springer-Verlag Berlin Heidelberg, 2007.
11. Ley, Michael. "The DBLP computer science bibliography: Evolution, research issues, perspectives." *International symposium on string processing and information retrieval*. Springer Berlin Heidelberg, 2002.
12. Aggarwal, Charu C., Yuchen Zhao, and S. Yu Philip. "Outlier detection in graph streams." *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*. IEEE, 2011.
13. Aggarwal, Charu C., and Jiawei Han, eds. *Frequent pattern mining*. Springer, 2014.
14. Cook, Diane J., and Lawrence B. Holder, eds. *Mining graph data*. John Wiley & Sons, 2006.
15. Dong, Guozhu, et al. "CAEP: Classification by aggregating emerging patterns." *International Conference on Discovery Science*. Springer Berlin Heidelberg, 1999.
16. Dong, Guozhu, and Jinyan Li. "Efficient mining of emerging patterns: Discovering trends and differences." *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1999.
17. Calders, Toon, Jan Ramon, and Dries Van Dyck. "Anti-monotonic overlap-graph support measures." *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, 2008.
18. Han, Jiawei, et al. "Frequent pattern mining: current status and future directions." *Data Mining and Knowledge Discovery* 15.1 (2007): 55-86.
19. Chan, Jeffrey, James Bailey, and Christopher Leckie. "Discovering correlated spatio-temporal changes in evolving graphs." *Knowledge and Information Systems* 16.1 (2008): 53-96.
20. Mongiovi, Misael, et al. "Netspot: Spotting significant anomalous regions on dynamic networks." *Proceedings of the 2013 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2013.
21. Vázquez, Alexei, et al. "Modeling of protein interaction networks." *Complexus* 1.1 (2002): 38-44.

22. Solé, Ricard V., et al. "A model of large-scale proteome evolution." *Advances in Complex Systems* 5.01 (2002): 43-54.
23. Beyer, Antje, et al. "Mechanistic insights into metabolic disturbance during type-2 diabetes and obesity using qualitative networks." *Transactions on Computational Systems Biology XII*. Springer Berlin Heidelberg, 2010. 146-162.
24. Liu, Siyuan, et al. "Mobile phone graph evolution: Findings, model and interpretation." *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*. IEEE, 2011.