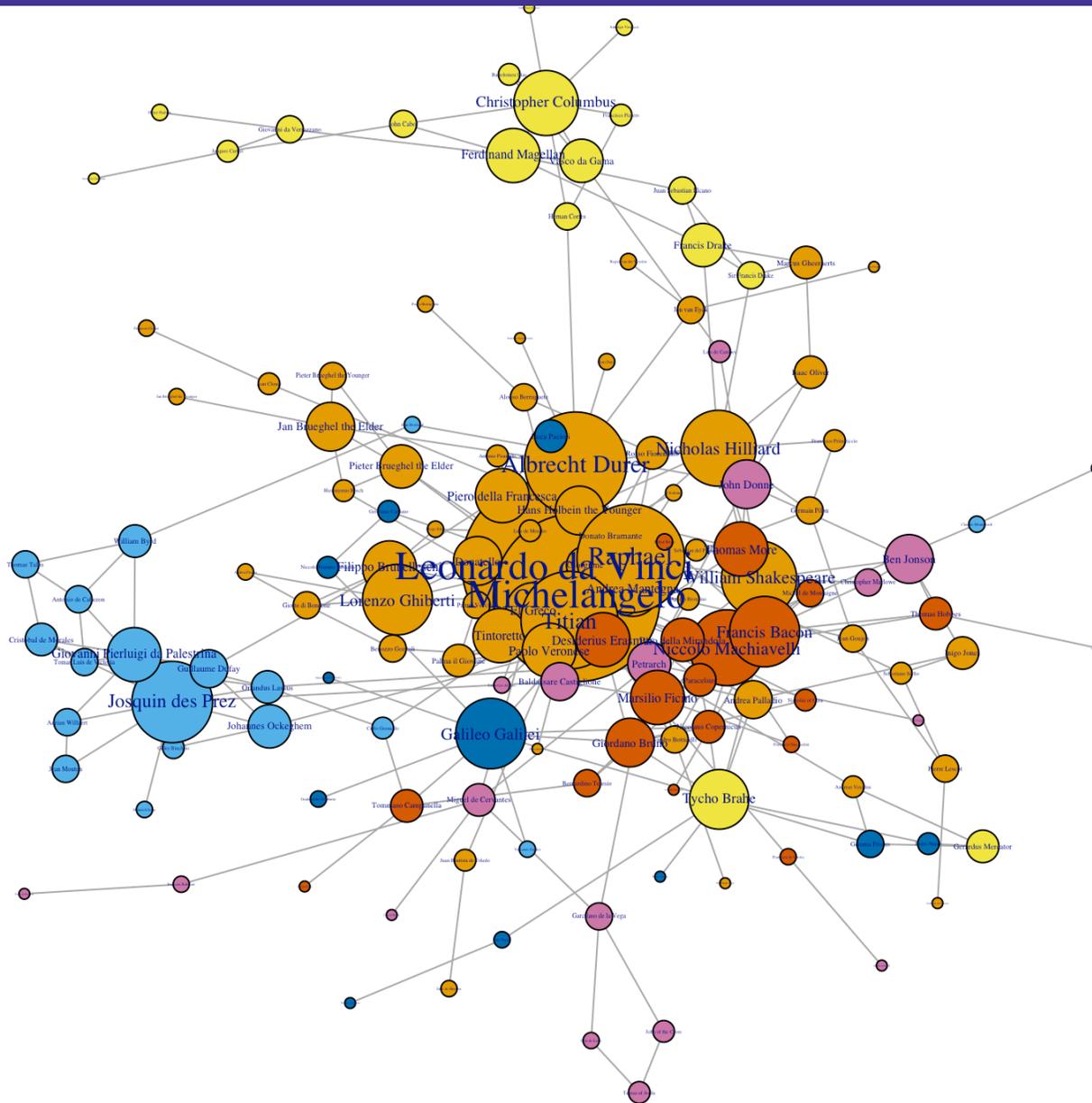
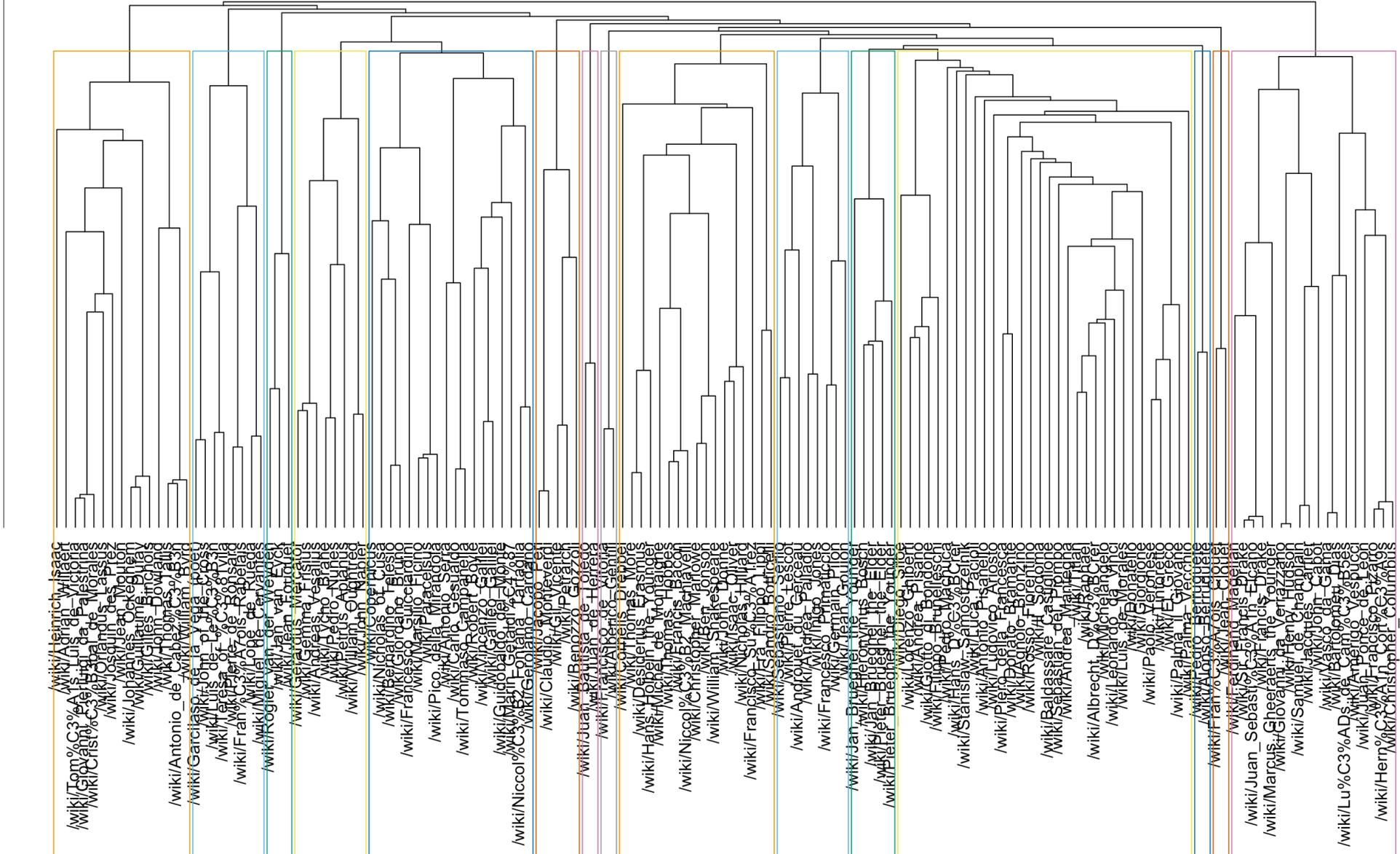


Esempio: personaggi del Rinascimento



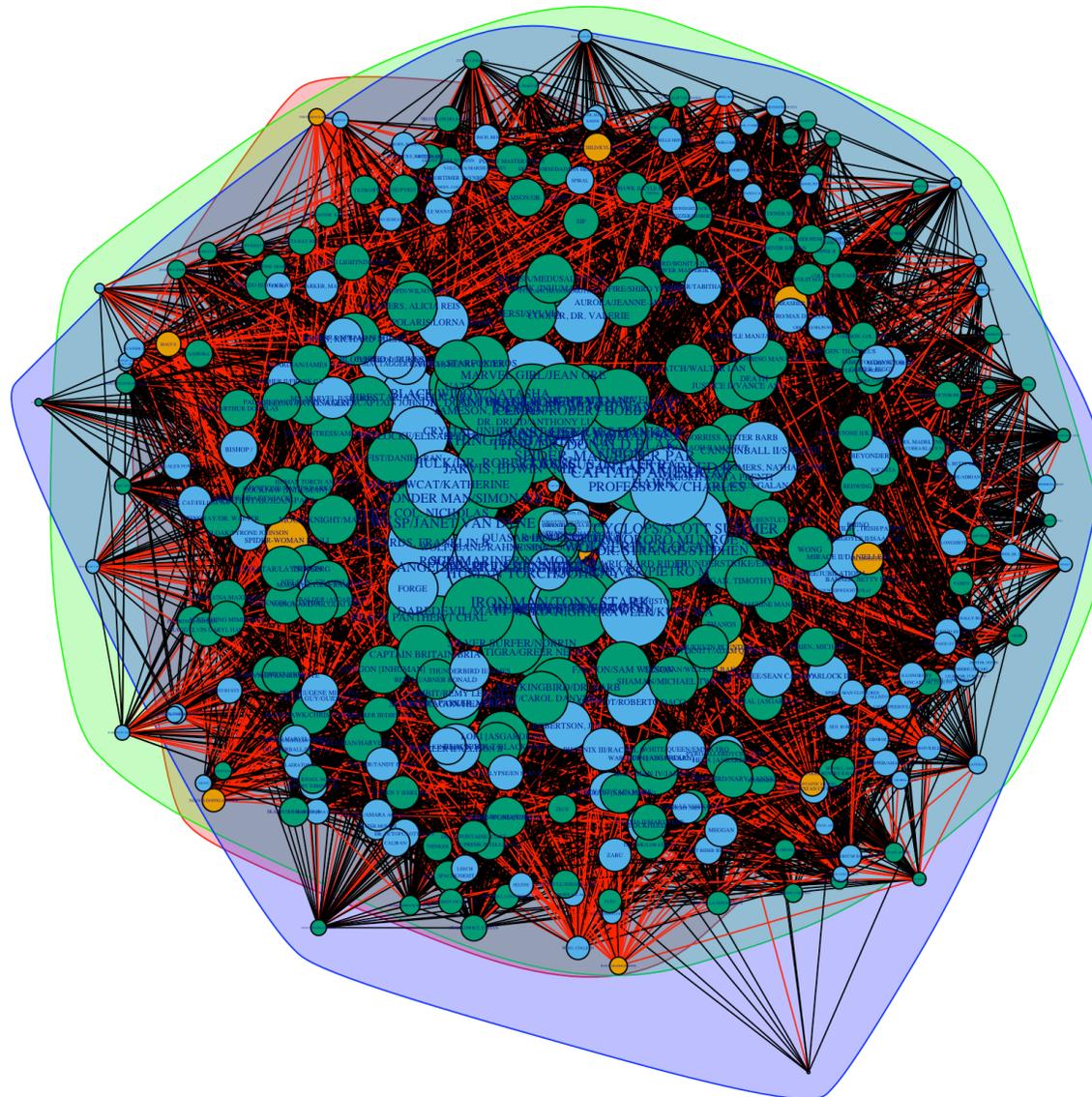
Esempio: personaggi del Rinascimento



Esempio: personaggi del Rinascimento

	degree	betweenness	closeness	eigenvector	subgraph
Michelangelo	28	2683.7946352	0.002710027	1.000000e+00	1308.625652
Leonardo_da_Vinci	23	1211.7120573	0.002538071	9.168138e-01	1099.311873
Raphael	15	824.6052928	0.002531646	6.463982e-01	551.237267
Titian	13	519.1673671	0.002347418	6.272625e-01	523.023811
Albrecht_D%BCr%BCrer	15	2139.7179806	0.002525253	5.704162e-01	434.433917
Andrea_Mantegna	8	294.7169877	0.002267574	4.139509e-01	228.429923
Paolo_Veronese	7	69.5149881	0.002217295	4.131263e-01	229.991870
Niccol%CCMachiavelli	12	697.6072463	0.002386635	4.030870e-01	230.863323
Tintoretto	6	27.6071915	0.002159827	3.961055e-01	211.159906
El_Greco	7	494.0357422	0.002283105	3.771884e-01	192.460372
Donatello	6	18.9726428	0.002109705	3.578871e-01	174.314671
Pico_della_Mirandola	6	103.0299815	0.002150538	3.143099e-01	136.134729
Piero_della_Francesca	7	80.0146464	0.002079002	3.111377e-01	135.295153
Lorenzo_Ghiberti	7	306.1085007	0.002145923	3.006118e-01	127.971474
Giorgione	5	10.3931541	0.002000000	2.976080e-01	119.959094
Luis_de_Morales	3	0.0000000	0.002074689	2.882223e-01	109.902512
Hans_Holbein_the_Younger	7	433.6711835	0.002257336	2.876777e-01	114.897718
Donato_Bramante	6	251.2106470	0.002105263	2.772080e-01	104.754368
Marsilio_Ficino	7	195.2664938	0.002202643	2.579350e-01	98.281584
Baldassare_Castiglione	5	636.9289357	0.002207506	2.551038e-01	90.596043
Filippo_Brunelleschi	7	312.9519978	0.002109705	2.402400e-01	85.896353
Pieter_Brueghel_the_Elder	5	202.4681818	0.002092050	2.380730e-01	78.662998
Palma_Vecchio	4	14.2576299	0.002000000	2.376109e-01	78.073321
Sebastian_del_Piombo	3	3.4047683	0.002032520	2.185952e-01	65.104722
Sandro_Botticelli	4	147.5357864	0.002083333	2.069347e-01	59.741422

Esempio: super-eroi Marvel



Esempio: super-eroi Marvel

	degree	betweenness	closeness	eigenvector	subgraph
CAPTAIN AMERICA	361	1115.17204	0.002544529	1.0000000	1.815272e+74
MR. FANTASTIC/REED R	346	885.75791	0.002450980	0.9752055	1.726370e+74
WOLVERINE/LOGAN	343	842.56530	0.002433090	0.9725709	1.717055e+74
SPIDER-MAN/PETER PAR	341	846.54561	0.002421308	0.9664849	1.695633e+74
IRON MAN/TONY STARK	339	787.11841	0.002409639	0.9660916	1.694253e+74
THING/BENJAMIN J. GR	336	777.89183	0.002392344	0.9595759	1.671477e+74
THOR/DR. DONALD BLAK	336	804.31846	0.002392344	0.9538079	1.651443e+74
BEAST/HENRY & HANK & P	327	703.05439	0.002341920	0.9459335	1.624288e+74
HUMAN TORCH/JOHNNY S	328	707.39010	0.002347418	0.9458470	1.623991e+74
CYCLOPS/SCOTT SUMMER	324	645.67224	0.002325581	0.9431692	1.614808e+74
SCARLET WITCH/WANDA	322	661.49421	0.002314815	0.9383031	1.598189e+74
INVISIBLE WOMAN/SUE	320	635.17187	0.002304147	0.9321818	1.577404e+74
VISION	317	607.90796	0.002288330	0.9295588	1.568540e+74
WASP/JANET VAN DYNE	314	570.98883	0.002272727	0.9271973	1.560580e+74
SHE-HULK/JENNIFER WA	316	617.42525	0.002283105	0.9265508	1.558404e+74
STORM/ORORO MUNROE S	311	578.84855	0.002257336	0.9182831	1.530717e+74
COLOSSUS II/PETER RA	307	550.11739	0.002237136	0.9102813	1.504156e+74
ANGEL/WARREN KENNETH	307	550.51620	0.002237136	0.9087801	1.499199e+74
HULK/DR. ROBERT BRUC	302	530.49848	0.002212389	0.8994602	1.468607e+74
HAWK	300	487.25325	0.002202643	0.8986018	1.465805e+74
HERCULES [GREEK GOD]	298	459.17296	0.002192982	0.8984058	1.465166e+74
PROFESSOR X/CHARLES	301	535.35420	0.002207506	0.8888657	1.434214e+74
WONDER MAN/SIMON WIL	291	426.92129	0.002159827	0.8873654	1.429377e+74
DR. STRANGE/STEPHEN	298	562.05938	0.002192982	0.8838317	1.418015e+74
ICEMAN/ROBERT BOBBY	296	512.96077	0.002183406	0.8807592	1.408173e+74

Modelli per grafi e network

Oltre a metodi per costruire grafi ed analizzare la loro struttura, è possibile individuare modelli e «classi» di network.

Per «modello» di network si intende l'insieme:

$$\{P_{\theta}(G), G \in \mathcal{G} : \theta \in \Theta\},$$

Dove \mathcal{G} è un insieme di possibili grafi, P_{θ} è una distribuzione di probabilità in \mathcal{G} , e θ è un vettore di parametri, con possibili valori in Θ .

La complessità di questi modelli deriva sostanzialmente dal modo in cui viene specificata la distribuzione di probabilità $P(\cdot)$. Alcuni approcci, suppongono per comodità, $P(\cdot)$ uniforme in \mathcal{G} .

Nella pratica, questi modelli sono usati oltre che per fini statistici (stime e test di significatività circa determinate caratteristiche), per l'individuazione di certe proprietà comuni (distribuzione dei gradi, centralità ecc)

Principali modelli

Le principali categorie di modelli sono:

1. Random Graph models
2. Small-world models
3. Network Growth models
4. Exponential Random Graph models

L'idea di fondo dei Random Graph model, è che la loro rappresentazione sia estratta casualmente dall'insieme di tutti i possibili grafi, specificando un determinato insieme G ed una distribuzione di probabilità uniforme $P(\cdot)$ in G .

«Classical» e «Generalized» Random Graph

La teoria dei Random Graph si divide in approccio classico e generalizzato: nel primo, il modello pone probabilità uguali per tutti i grafi di determinati ordine e grandezza.

In formule, viene specificato un insieme G_{N_v, N_e} di tutti i grafi $G = (V, E)$ con $|V| = N_v$ ed $|E| = N_e$, ed assegnata una probabilità $P(G) = \frac{1}{\binom{N}{N_e}}$ ad ogni $G \in G_{N_v, N_e}$, dove $N = \binom{N_v}{2}$ è il numero totale di coppie di vertici.

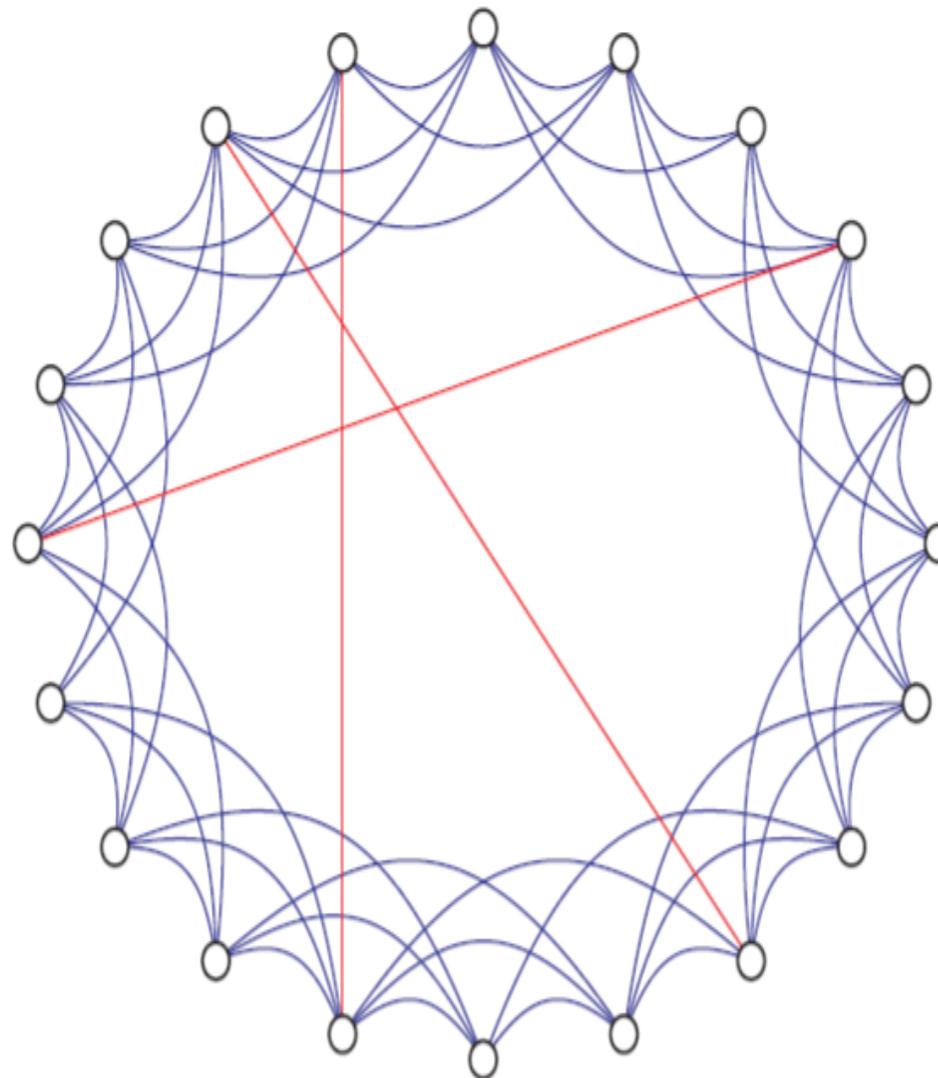
L'approccio generalizzato, invece, si sostanzia nel:

- Definire un insieme G di grafi di ordine N_v
- Assegnare uguale probabilità ad ognuno dei grafi $G \in G$.

Small-World Models

I modelli di questo tipo hanno avuto una sempre più crescente importanza, poiché sono espressamente costruiti per rappresentare alcune proprietà osservate nel *mondo reale*.

Un network può essere definito Small World quando la sua struttura si compone di un sistema di scambi cosiddetto *neighbor-to-neighbor* (tra vicini), sufficiente per la trasmissione di informazioni a tutta la rete.



Network Growth Models

I cosiddetti modelli di crescita del Network hanno ragione di esistere poiché è comune osservare reti che evolvono (e spesso crescono) nel tempo, come ad esempio la rete internet globale (World Wide Web), o un network composto da citazioni scientifiche.

L'obiettivo dell'analisi di questi modelli è quello di osservare le proprietà che emergono dalla rete considerando un ampio numero di successivi periodi temporali, ed in questo contesto, due sono i principali meccanismi di crescita, denominati:

Network Growth Models

- **Preferential Attachment:** questo meccanismo incorpora il principio secondo il quale «il ricco diventa sempre più ricco» e rappresenta crescite di network notevoli nel tempo (es. la crescita di popolarità di determinati siti o App) ed uno dei modelli principali è stato proposto da Barabàsi-Albert per grafi non diretti.

Partendo da un grafo iniziale $G^{(0)}$ di $N_v^{(0)}$ vertici ed $N_e^{(0)}$ spigoli. Ad ogni stadio $t=1,2,\dots$, il Grafo corrente $G^{(t-1)}$ è modificato per creare un nuovo grafo $G^{(t)}$ aggiungendo un nuovo vertice di grado $m \geq 1$, dove gli m nuovi spigoli sono uniti ad m differenti vertici in $G^{(t-1)}$, e la probabilità che il nuovo vertice sia connesso ad un dato vertice v è data da

$$\frac{d_v}{\sum_{v \in V} d_v} \cdot$$

Dopo t iterazioni quindi, il grafo risultante $G^{(t)}$ avrà $N_v^{(t)} = N_v^{(0)} + t$ vertici e $N_e^{(t)} = N_e^{(0)} + tm$ spigoli.

Network Growth Models

- **Copying**: anch'esso presentato in relazione a strutture di network come il web ma anche in campo biochimico, incorpora il principio della «duplicazione».

Partendo da un grafo iniziale $G^{(0)}$, nuovi grafi $G^{(t)}$ sono costruiti immediatamente dal loro predecessore $G^{(t-1)}$, aggiungendo un nuovo vertice v connesso in maniera casuale ad un subset di vertici adiacenti ad un vertice u .

Più precisamente, un vertice u è scelto casualmente da $G^{(t-1)}$ ed il nuovo vertice v sarà collegato a tutti i vicini di u in $G^{(t-1)}$ con una certa probabilità p .

Inferenza sulla topologia del Network

I problemi di inferenza sulla topologia del Network, si sostanziano nel valutare, tra tutti i possibili grafi G contenuti in un set \mathcal{G} , quale meglio cattura la struttura sottostante il sistema, basandosi su:

1. *Informazioni contenute nei dati raccolti:* ad esempio un set \mathbf{x} di attributi riguardanti i vertici, $\mathbf{x} = (x_1, \dots, x_{N_v})^T$;
2. *Qualsiasi informazione a priori sulla rete:* ad esempio un set di indicatori binari $\mathbf{y} = [y_{ij}]$ riguardanti elementi del network (come la presenza di connessione o meno fra due vertici).

Tre tipi di analisi

- *Link Prediction*: Analisi inferenziale volta a valutare se una coppia di vertici è connessa (inferenza *edge* o *non-edge*), usando misurazioni che includono coppie di vertici il cui stato di connessione o meno è già noto.
- *Association Networks*: Analisi inferenziale volta alla definizione di associazioni (link) tra i vertici che non siano note a priori, utilizzando per l'inferenza caratteristiche note dei vertici stessi.

N.B: in questo tipo di analisi non si considerano livelli di associazione banali (come la correlazione)
- *Tomographic Network inference*: Questa analisi si distingue dalle altre per il fatto che le misurazioni sono disponibili solo per i vertici che giacciono sul «perimetro» del network, ed è volta all'individuazione sia di vertici che di link all' «interno» dello stesso.

Relazione fra le diverse analisi

L'ordine di queste tre tipi di problemi segue una progressione:

Il primo assume la conoscenza di caratteristiche di tutti i vertici e per alcuni anche dello stato *edge-non edge*, e cerca di inferire il resto di questi ultimi. Nel secondo non v'è conoscenza di link (*edge status*) in nessuna parte del network, ma si assumono note importanti informazioni dei vertici le quali vengono utilizzate per l'inferenza sulle relazioni fra questi ultimi.

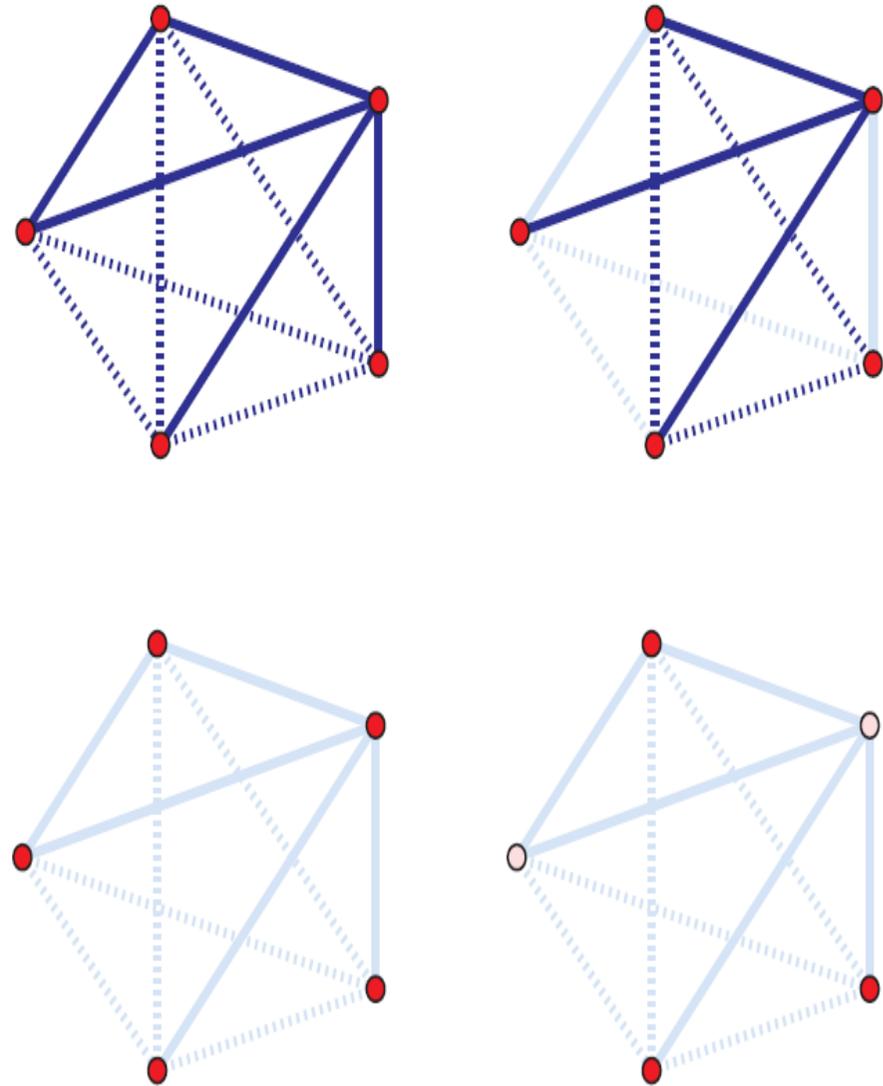
Il terzo, infine, implica misurazioni note solo per un particolare subset di vertici, che però forniscono informazioni utili per l'inferenza sulla topologia della parte non nota del network.

Esempio dei tre tipi di inferenza sulla topologia, per un grafo G

Esempio dei tre tipi di inferenza sulla topologia, per un grafo G.

Gli «edge» sono mostrati con la linea retta, i «non edge» con quella tratteggiata. I vertici e gli edge osservati sono mostrati con colore scuro (rispettivamente rosso e blu), mentre i vertici e gli edge non osservati hanno colori chiari (rosa e celeste).

In alto a sinistra il Grafo reale sottostante, in alto a destra l'analisi *link prediction*, in basso a sinistra l'*Association graph inference*, mentre in basso a destra, *Tomographic network inference*.



Relazione fra le diverse analisi

Spesso, nella conduzione di un'analisi di un network, oltre alle interazioni tra gli elementi oggetto di studio è possibile dover esaminare una qualche caratteristica o quantità a loro associata, la quale può essere influenzata e variare in relazione alle diverse interazioni presenti nel network.

Ad esempio, i comportamenti o idee degli individui variano a seconda delle loro relazioni sociali; computer facilmente accessibili da un computer infetto da virus, possono diventare loro stessi infetti più velocemente, ecc..

Quantità associate a questo tipo di fenomeni possono essere concepite come processi stocastici, definiti all'interno del network.

Processi statici e dinamici

Formalmente, suddette quantità possono essere rappresentate in termini di sequenze di variabili casuali, \mathbf{X} , indicizzate totalmente o in parte al grafo $G = (V, E)$, nella forma $\{X_i\}$, per $i \in V$, o $\{X_i(t)\}$, per $i \in V$ e $t \in T$, dove T è un range di periodi temporali (discreto o continuo).

Quindi, diversi comportamenti degli individui possono essere rappresentati con variabili categoriali associate ad ogni $i \in V$. Analogamente, la diffusione di un virus informatico può essere rappresentata *con* set di variabili binarie (*infetto/non infetto*)

Ci si riferisce, con processi del tipo $\{X_i\}$, a *processi statici*, mentre con $\{X_i(t)\}$ a *processi dinamici*.

Nearest Neighbor Prediction

È un metodo per la previsione di un processo statico.

Consideriamo un insieme di caratteristiche dei vertici $\mathbf{X} = (X_i)_{i \in V}$.

Tali attributi, possono essere indipendenti da t , oppure costituire una «istantanea» di un processo dinamico in un dato momento temporale.

Abbiamo visto precedentemente come tali informazioni possano contribuire alla previsione di presenza o assenza di relazioni nel grafo, modellando quindi il comportamento di $\mathbf{Y} = [Y_{i,j}]_{(i,j) \in V^{(2)}}$, condizionatamente a \mathbf{X} .

Il metodo del «vicino più prossimo» (*Nearest Neighbor*) utilizza la media di attributi:

Nearest Neighbor Prediction

$$\frac{\sum_{j \in N_i} X_j}{|N_i|}$$

osservati nei vertici prossimi ad un vertice i , per effettuare previsioni circa la realizzazione (ignota) di quell'attributo in i .

È il concetto usato in maniera più diffusa, poiché conoscere una caratteristica di elementi prossimi ad un elemento i , in un network, è un'ottimo modo per effettuare inferenza sulla stessa caratteristica in i .

Un problema può sorgere se alcuni delle X_j nell'equazione precedente siano ignote: in tal caso, o si cerca di inserire nell'equazione solo i vertici adiacenti e con informazione nota, o si usano metodi di imputazione per le X_j mancanti (ad esempio inserendo la media generale delle osservazioni note).

Markov Random Fields

Dati un grafo $G = (V, E)$ e un set $\mathbf{X} = (X_1, \dots, X_{N_v})^T$ di variabili casuali discrete definite in V , si dice che \mathbf{X} è un *Markov random field (MRF)* in G se:

$P(\mathbf{X} = \mathbf{x}) > 0$, per ogni possibile outcome \mathbf{x} ,

e

$$P(X_i = x_i \mid \mathbf{X}_{(-i)} = \mathbf{x}_{(-i)}) = P(X_i = x_i \mid \mathbf{X}_{N_i} = \mathbf{x}_{N_i}),$$

Dove $\mathbf{X}_{(-i)}$ è il vettore $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{N_v})^T$ e \mathbf{X}_{N_i} è il vettore delle X_j , per ogni $j \in N_i$.

La seconda condizione è detta «Markoviana», è afferma che X_i , è condizionatamente indipendente da tutte le altre X_k , dati i valori dei vertici adiacenti, definiti dalla struttura del grafo G .

Applicazioni del M.R.F.

Modelli sviluppati con il Markov Random Field possono essere utilizzati per la previsione di alcune o tutte le $(X_i)_{i \in V}$ in processi statici, e possono essere di vario tipo, in base alla forma dell'equazione «Markoviana» vista in precedenza, un esempio è:

- *Auto-Logistic MRF Model*, dove

$$P(X_i = 1 \mid \mathbf{X}_{N_i} = \mathbf{x}_{N_i}) = \frac{\exp(\alpha + \beta \sum_{j \in N_i} X_j)}{1 + \exp(\alpha + \beta \sum_{j \in N_i} X_j)};$$

Questo modello può essere esteso al caso in cui le X_j assumano valori $\{0, 1, \dots, m\}$ per un intero positivo m , creando in tal modo una classe di modelli chiamata *Multi-level logistic* o *multi-color Models*.

Regressioni Kernel

Per gli obiettivi previsionali, metodi di regressione standard come il classico metodo dei minimi quadrati che suppone la relazione tra le variabili risposta e quelle predittive definita nello spazio euclideo, non sono immediatamente applicabili a dati indicizzati nei grafi.

I cosiddetti «Metodi Kernel» si prestano in maniera migliore a questo tipo di dati, ed essi consistono in una nozione generalizzata di predittori (codificati in un cosiddetto «Kernel»)

Esempio di processo dinamico: il Modello epidemico

Molti sistemi studiati partendo da Network rappresentativi, concepiscono le relazioni in maniera dinamica, come la diffusione di informazioni, malattie, ecc..

Si può pensare a processi di questo tipo come a processi di attributi di vertici indicizzati nel tempo, $X(t) = X_i(t)$, per $i \in V$ e $t \in T$, dove T è un range di tempi. Per modellare questo tipo di processi vengono utilizzati sia prospettive deterministiche che stocastiche, basate rispettivamente su equazioni differenziali e processi Markoviani.

Un termine usato per definire processi che si delineano nel tempo in maniera maggiore rispetto alle aspettative, è il termine *epidemico*.

Esempio di processo dinamico: il Modello epidemico

I modelli epidemici hanno a che fare con 3 tipi di problemi:

- Capire il meccanismo che causa la diffusione;
- Prevedere il percorso futuro dell' «epidemia»;
- Raggiungere un'abilità di controllo della diffusione.

La più comunemente usata classe di modelli epidemici continui nel tempo è detta *SIR (susceptible-infected-removed)* ed il membro più importante di tale classe è il *Modello epidemico generale*.

Il concetto alla base è che, dati $N+1$ elementi nel network, in ogni momento t ci sarà un numero casuale $N_S(t)$ suscettibile di infezione, $N_I(t)$ elementi infetti e $N_R(t)$ elementi immuni (o rimossi). In ogni tempo t un nuovo fra i suscettibili diverrà infetto con probabilità proporzionale al prodotto tra il numero dei suscettibili s e quello degli infetti i . Analogamente, gli infetti guariranno con probabilità proporzionale ad i .

Analisi dei network di flusso

Lo scopo di molteplici network è quello di essere «conduttori» di flussi, ovvero di facilitare il movimento di qualcosa, come materiali, persone, informazioni.

Esempi li troviamo nella vita di tutti i giorni:

- ❖ Rete stradale o ferroviaria → Flussi di persone, merci ...
- ❖ Reti di comunicazioni → Flussi di informazioni, dati ...
- ❖ Rete commerciale → Flusso di capitali, beni ...

I Grafi per flussi

Sia $G = (V, E)$ il grafo di un network. Dato che un qualsiasi flusso presenta una direzione, da un'origine ad una destinazione, formalmente G sarà un grafo diretto.

Il flusso attraversa molteplici link (e_1, \dots, e_n) nel passaggio dal vertice di origine a quello di destinazione nel network, e il modo in cui il traffico si sposta attraverso la rete è rappresentato da una matrice, \mathbf{B} , chiamata *routing matrix* (*matrice di instradamento*), avente numero di righe pari ai link attraversati, e numero di colonne pari alle coppie di *origine-destinazione* in G .

Matrice binaria \mathbf{B}

Nel caso in cui, ogni coppia di origine-destinazione (i,j) sia unita da un singolo link, da i a j , \mathbf{B} sarà una matrice binaria, composta da:

$$B_{e;ij} = \begin{cases} 1, & \text{se il link } e \text{ è attraversato, andando da } i \text{ a } j \\ 0, & \text{altrimenti} \end{cases}$$

\mathbf{B} descrive quindi l'incidenza dei percorsi con i link. Se sono presenti percorsi multipli (più lunghi di un singolo link), i dati in \mathbf{B} saranno frazioni rappresentanti la porzione del traffico da i a j , spettante al link e .

	0	1	0	...
	0	0	1	...
	1	0	0	...
...

Matrice Origine-Destinazione

Nell'analisi dei flussi, è di fondamentale importanza una quantità chiamata *matrice origine-destinazione (OD)*, denominata $\mathbf{Z} = [Z_{ij}]$, dove Z_{ij} è il volume totale del traffico che viaggia da un vertice d'origine i ad uno di destinazione j in un dato periodo di tempo. Ci si riferisce a tale matrice anche con l'espressione: «*matrice di traffico*».

Altre quantità d'interesse sono relazionate a \mathbf{Z} : ad esempio, il flusso netto in entrata ed in uscita, corrispondente ai vertici i e j può essere rappresentato, rispettivamente, come:

$$Z_{i+} = \sum_j Z_{ij} \quad \text{e} \quad Z_{j+} = \sum_i Z_{ij}$$

Matrice Origine-Destinazione

Inoltre, se definiamo X_e come il flusso totale su un dato link $e \in E$, e $\mathbf{X} = (X_e)_{e \in E}$, il totale dei link in \mathbf{X} può essere relazionato a \mathbf{Z} con l'espressione:

$$\mathbf{X} = \mathbf{BZ}$$

Dove \mathbf{Z} ora rappresenta la matrice di traffico, espressa in forma vettoriale.

Oltre alle misure di volume di traffico, sono importanti due nozioni:

- **Costo (c):** spesso associato a percorsi o singoli link, è molto utilizzato in campo socioeconomico e nell'analisi dei trasporti, per modellare il comportamento dei consumatori (attori del network) nelle loro scelte;
- **Qualità del servizio:** è utilizzata nella letteratura sulle reti e sul traffico informatico, ove è di comune interesse per tutti gli attori (provider e consumatori) monitorare le performance del network, e identificare eventuali aree di malfunzionamento.

Modelli gravitazionali

I modelli gravitazionali sono una classe di modelli, sviluppata largamente nelle scienze sociali, per descrivere livelli aggregati di interazione tra soggetti di diverse popolazioni. Intuitivamente, il loro utilizzo spazia dalla geografia, all'economia, alla sociologia ma le applicazioni possono riguardare anche campi scientifici.

Il termine «*modello gravitazionale*» deriva dal fatto che, analogamente alla legge di Newton sull'attrazione gravitazionale, l'interazione tra due popolazioni varia in maniera direttamente proporzionale alla loro grandezza, ed inversamente ad alcune misure della loro separazione.

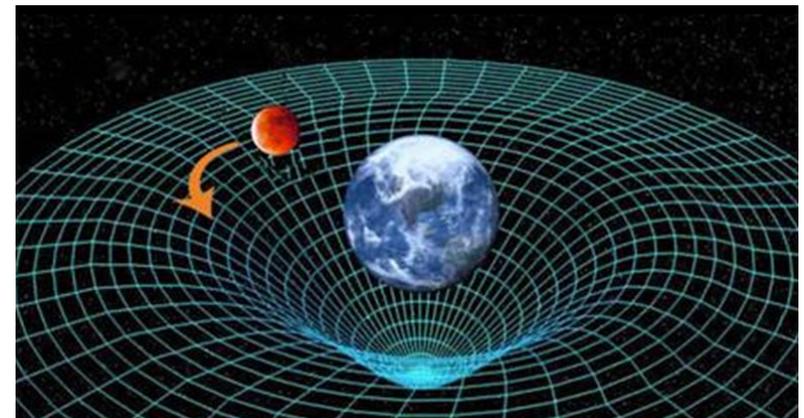
Modello gravitazionale generale

Supponiamo che I e J rappresentino set di origini e destinazioni, di cardinalità $|I|$ e $|J|$ rispettivamente, e Z_{ij} denoti una misura di traffico fluente da $i \in I$ a $j \in J$ in un dato periodo di tempo.

Il **modello gravitazionale generale** specifica che il flusso di traffico Z_{ij} è, sottoforma di conteggi, formato da distribuzioni di Poisson indipendenti, con media:

$$E(Z_{ij}) = h_O(i)h_D(j)h_S(\mathbf{c}_{ij})$$

Dove h_O, h_D sono funzioni positive rispettivamente dell'origine i e della destinazione j , e h_S è funzione di un vettore di attributi di separazione \mathbf{c}_{ij} .



Modello gravitazionale generale

I K elementi all'interno del vettore c_{ij} servono a quantificare la separazione tra la coppia di vettori di origine e destinazione (i,j) e spesso sono basati sul concetto di «distanza» o di «costo».

Le funzioni precedenti, h_O e h_D , sono chiamate anche *funzione di origine* e *funzione di destinazione*, mentre h_S è chiamata *funzione di separazione* o *funzione di deterrenza* (poiché la distanza può essere considerata come una forma di deterrenza, che può influire in maniera inversamente proporzionale sulla quantità dei flussi nel network).

Modello di gravitazione demografica

Un più recente esempio di modello gravitazionale, proposto da Stewart all'interno della sua teoria della «*Gravitazione demografica*», specifica che:

$$E(Z_{ij}) = \gamma \pi_{O,i} \pi_{D,j} d_{ij}^{-2}$$

Dove $\pi_{O,i}$ e $\pi_{D,j}$ sono misure della grandezza della popolazione di due regioni geografiche di origine e destinazione, rispettivamente, mentre $d_{i,j}$ è una misura della distanza dei centri delle due regioni. Questa formulazione è analoga alla Legge Universale di Newton, ad eccezione dell'uso di una «costante di gravitazione demografica» γ .

La stima del flusso nel network

Il monitoraggio e la misurazione del volume di flusso Z_{ij} è un problema riguardante molteplici tipi di network. Ad esempio, nelle reti stradali, vengono usate numerose indagini che possono portare in accuratezza dei risultati a piccola scala a causa della numerosità campionaria contenuta e degli errori campionari. (*roadside surveys, home or destination surveys, ecc*).

Lo sviluppo tecnologico ha portato a sistemi e strumenti di monitoraggio automatico come sensori, telecamere ecc.. i quali riducono il problema grazie al loro impiego su larga scala, ma solo di recente si è riusciti ad inserire direttamente all'interno di alcuni network (come Internet) sistemi per immagazzinare e trasmettere informazioni sul volume dei flussi, in tempo reale.

La stima del flusso nel network

Tuttavia, la conoscenza del flusso di traffico Z_{ij} è fondamentale per svariati obiettivi come gestione del traffico, approvvigionamento della rete, manutenzione e pianificazione di crescita della stessa.

In molti contesti dove è difficile quantificare il flusso, è più facile misurare il volume X_e sui links del network come ad esempio, posizionando sensori all'inizio e alla fine di rampe o singoli tratti nelle reti autostradali.

In casi come questo, ci troviamo di fronte al problema della previsione di Z_{ij} , o alternativamente, della stima delle sue medie μ_{ij} , avendo osservato i conteggi dei links $\mathbf{X} = (X_e)_{e \in E}$.

Questo problema è denominato «*stima della matrice di flusso*»

Metodi per la stima della matrice di flusso

Il problema della quantificazione del flusso in un network può essere affrontato mediante due classi di metodi.

A seconda che l'obiettivo sia la stima della matrice di flusso per un singolo periodo di tempo, o per molteplici periodi successivi, questi metodi si dividono in:

- Metodi statici
- Metodi dinamici

Inoltre, nell'utilizzo di questi metodi si suppone che la matrice **B** precedentemente esaminata sia nota, poichè la sua costruzione è un obiettivo a sé stante.

Modelli gravitazionali per il Commercio

Questi modelli, chiamati «Gravity Models», per l'analisi del commercio internazionale, definiscono il flusso commerciale X tra due paesi come:

$$X_{ij} = G \frac{Y_i^{\beta_1} Y_j^{\beta_2}}{D_{ij}^{\beta_3}}$$

Dove Y_i e Y_j rappresentano le grandezze dei paesi esportatori e importatori, spesso espresse sottoforma di PIL (GDP). D rappresenta la loro distanza geografica mentre G è una costante non di fondamentale interesse. Secondo l'equazione, più è grande il PIL dei due paesi e più vicini sono tra di loro, più le loro relazioni commerciali saranno grandi.

I coefficienti β_i rappresentano l'elasticità e in questo modello, è spesso assunto che l'elasticità dei redditi siano uguali a uno, ovvero $\beta_1 = \beta_2 = 1$.

Elementi di variazione del Modello

All'interno del modello, possono essere inseriti determinate componenti per rappresentare in maniera migliore le relazioni fra i paesi, come ad esempio:

- Un termine t_{ij} rappresentante il costo delle transazioni tra i due paesi, contenente variabili dummy che indichino confini condivisi (contig) o accordi siglati in termini commerciali (rta):

$$t_{ij} = D_{ij} e^{contig_{ij}} e^{rta_{ij}} = \exp(\log D_{ij} + contig_{ij} + rta_{ij})$$

Ed il modello può quindi essere riscritto in forma:

$$\log t_{ij} = \beta_0 \log G + \beta_1 \log Y_i + \beta_2 \log Y_j + \beta_3 \log D_{ij} + \beta_4 contig_{ij} + \beta_5 rta_{ij}$$

Elementi di variazione del Modello

Ed in questa forma additiva, il modello può essere stimato con metodi standard come l'OLS.

Ciò che il modello assume però, sono solo costi bilaterali fra due paesi, ignorando quelli col resto del mondo; e per poter considerare anche gli effetti che altri paesi possono avere su un commercio bilaterale, viene inclusa una misura di «*Resistenza Multilaterale*», (MR).

Le quantità di importazioni ed esportazioni tra i e j , sono influenzate da altre domande per i beni di i e altre offerte per i beni di j ed inoltre, il volume di commercio tra due paesi può decrescere come conseguenza della variazione delle barriere commerciali tra i due paesi e tutti gli altri.

Alcune conclusioni di vari autori affermano che piccole nazioni reagiscono maggiormente a variazioni delle barriere commerciali rispetto a grandi nazioni.

Modello gravitazionale simmetrico

Ed il modello, col termine MR al suo interno sarà:

$$X_{ij} = \frac{Y_i Y_j}{Y} \frac{t_{ij}^{1-\sigma}}{P_j^{1-\sigma} P_i^{1-\sigma}}$$

Dove il flusso X_{ij} da i a j è determinato dai fattori dell'offerta in i , Y_i , dai fattori della domanda in j , Y_j , dai costi transazionali t_{ij} da i a j (assunti simmetrici, ovvero $t_{ij} = t_{ji}$).

$Y = \sum_i Y_i = \sum_j Y_j$ rappresenta la somma dei redditi mondiali, mentre σ è l'elasticità di sostituzione tra tutti i beni ed è assunta $\sigma > 1$ per preferire la variazione degli stessi. Il termine MR è incluso con il termine P, funzione dei costi di transazione di i con tutti i partner commerciali j .

Metodi di stima del modello

Nonostante l'idea di fondo di tali modelli sia semplice, la complessità risiede nella scelta dei metodi di stima. Possono essere divisi in due macro aree, la prima con metodi che stimano i modelli in forma logaritmica:

- *Ordinary Least Squares (OLS), Fixed Effects, Double Demeaning (DDM), Bonus vetus OLS with simple averages (BVU), Bonus vetus OLS with GDP weights (BVW), Structural Iterated Least Squares (SILS), Tetrads e alcuni metodi Tobit.*

E la seconda con metodi che stimano il modello nella forma moltiplicativa:

- *Poisson Maximum Likelihood (PPML), Gamma Pseudo Poisson Maximum Likelihood (GPML), Negative Binomial Pseudo Maximum Likelihood (NBPML), Nonlinear Least Squares (NLS)*