Reti come grafi dalla struttura disomogenea

1 Introduzione

Il materiale descritto in queste lezioni è contenuto nel capitolo 3 del testo e nella dispensa “Communities” [lucidi 12-29].

Fino ad ora abbiamo studiato grafi che modellano reti a livello statistico, utilizzando pesantemente il calcolo delle probabilità. In questo modo, ci siamo disinteressati delle peculiarità dei singoli nodi; piuttosto, abbiamo studiato proprietà che valevano mediamente per i nodi.

Cominciamo ora una serie di lezioni in cui cambiamo sostanzialmente il punto di vista. Analizziamo la posizione dei singoli nodi all'interno della rete per studiare:

- se è possibile evidenziare particolari strutture nella rete;
- se tutti i nodi si trovano, grosso modo, nella stessa situazione all'interno della rete, oppure vi sono delle differenze osservabili fra nodo e nodo;
- e se i nodi possono trarre vantaggio (e di quale tipo) da queste differenze.

2 L'esperimento di Granovetter

Negli anni '60 il sociologo statunitense Mark Granovetter, nel preparare la sua tesi di dottorato, intervistò un gruppo di individui che avevano recentemente cambiato lavoro. Fece loro una serie di domande volte a capire in che modo erano venuti a conoscenza della possibilità di ottenere l'impiego che avevano ottenuto (come avevano saputo dell'esistenza dell'azienda, che l'azienda cercava personale, ecc.).

Il risultato fu che molti degli intervistati avevano avuto le informazioni che li avevano condotti alla loro occupazione attuale attraverso una comunicazione da parte di qualcuno che conoscevano, una sorta di passaparola. Questo può ben essere prevedibile.

Il fatto abbastanza inaspettato che emerse dalle sue interviste fu che spesso l'informazione era arrivata da qualcuno che conoscevano superficialmente - non dagli amici più stretti. Questo è sorprendente, perché uno si aspetta che i nostri più cari amici sono quelli che più si darebbero da fare per aiutarci nella ricerca di un lavoro.

Allora, l'esperimento di Granovetter sembra suggerire che non è tanto la forza del legame quella che ha aiutato a trovare lavoro, quanto il tipo di informazione che una relazione è capace di veicolare. In effetti, pensandoci bene, i nostri più cari amici frequentano, grosso modo, gli stessi posti che frequentiamo noi e, così, abbiamo accesso alle stesse fonti di informazione. Invece, i nostri conoscenti (alla lontana) frequentano posti diversi, cosicché hanno accesso a fonti di informazione diverse, provenienti da "zone" diverse della rete, alle quali non avremmo accesso senza il loro aiuto e, così, ampliano il nostro "raggio d'azione".

3 Bridges e Local Bridges

Se consideriamo il grafo $G = (V, E)$ che modella la rete, gli archi che "aumentano" le nostre informazioni sono quelli che ci collegano a regioni della rete inaccessibili ai nostri amici.

Definizione 1 (Bridge)**.** Un **bridge** è un arco la cui rimozione disconnette la rete.

Segue dall'esperimento di Granovetter che i bridge sono gli archi che hanno maggiore "valore informativo". D'altra parte, sappiamo che una rete contiene con buona probabilità componenti giganti densamente connesse e, dunque, la presenza di bridge è poco probabile.

Pertanto, rilassiamo la definizione:

Definizione 2 (Local Bridge)**.** Un arco $(u, v) \in E$ è un **local bridge** se i suoi estremi non hanno vicini in comune, ossia se $N(u) \cap N(v) = \emptyset$, dove $N(u)$ è l'insieme dei nodi adiacenti al nodo u.

4 Legami Forti e Deboli

L'esperimento di Granovetter ci fornisce anche un altro tipo di indicazione. Poiché spesso l'informazione era arrivata da qualcuno conosciuto solo superficialmente, possiamo riconoscere all'interno di una rete sociale:

- **Relazioni forti**: quelle che ci collegano agli amici stretti.
- **Relazioni deboli**: quelle che ci collegano ai semplici conoscenti.

Conseguentemente, possiamo modellare una rete di questo tipo mediante un grafo G in cui gli archi sono partizionati in due sottoinsiemi: archi forti (strong ties) e archi deboli (weak ties), ossia $G = (V, S \cup W)$ con $S \cap W = \emptyset$.

Come ha mostrato l'esperimento, gli archi deboli veicolano informazioni alle quali non avremmo accesso tramite gli archi forti: e questa è la **forza degli archi deboli**.

4.1 Legami Forti e Chiusura Triadica

I nostri più cari amici frequentano, grosso modo, gli stessi posti che frequentiamo noi, e così hanno buone probabilità di entrare in contatto fra loro. Perciò, se c'è una relazione di amicizia forte fra A e B, e una tra A e C (cioè $(a, b) \in S$ e $(a, c) \in S$), allora è probabile che prima o poi si creerà anche l'arco (b, c), generando una **chiusura triadica**.

Alla base della creazione di chiusure triadiche possiamo individuare ragioni di:

- **Opportunità** di incontrarsi.
- **Incentivo** a stringere una relazione (se voglio andare al cinema con il mio amico A e lui ci va sempre con il suo amico B, mi conviene diventare amico di B).
- **Fiducia** (se mi fido di A che si fida di B, posso fidarmi anch'io di B).

4.2 Stabilità e Proprietà della Chiusura Triadica Forte (STCP)

Lasciamo che la rete evolva, formando triangoli, fino a raggiungere una configurazione stabile in cui non cambia più nel tempo. In questa configurazione, possiamo pensare che tutte le coppie di nodi collegate da un arco forte a uno stesso nodo siano fra loro collegate.

Definizione 3 (Strong Triadic Closure Property - STCP)**.** Sia $G = (V, S \cup W)$. Un nodo $u \in V$ soddisfa la **proprietà della chiusura triadica forte (STCP)** se, per ogni coppia di archi forti incidenti su u, i loro estremi sono collegati da un arco:

$$\forall (u, v) \in S, \forall (u, z) \in S \implies (v, z) \in S \cup W$$

Il grafo G soddisfa la STCP se tutti i suoi nodi la soddisfano.

Teorema 1. *Sia $G = (V, S \cup W)$. Se un nodo $u \in V$ soddisfa la STCP e se esiste un nodo x tale che $(u, x) \in S$, allora ogni local bridge (u, z) incidente su u deve essere un arco debole, cioè $(u, z) \in W$.*

Proof. Poiché (u, z) è un local bridge, allora $N(u) \cap N(z) = \emptyset$. Se per assurdo fosse $(u, z) \in S$, allora, poiché anche $(u, x) \in S$ e u soddisfa la STCP, dovrebbe essere $(x, z) \in S \cup W$. Ma questo implicherebbe che $x \in N(z)$, e dato che x è anche un vicino di u, si avrebbe $x \in N(u) \cap N(z)$, contraddicendo l'ipotesi che $N(u) \cap N(z) = \emptyset$. □

Questo teorema mostra come una proprietà globale (essere un local bridge) si rifletta in una proprietà locale (essere un weak tie).

5 Comunità e Clustering

Una chiusura triadica dopo l'altra, si formerà nella rete un gruppo di nodi fortemente coeso, con un elevato grado di interconnessione, che chiamiamo **cluster** o **comunità**. Ma quando possiamo dire che un gruppo di nodi è abbastanza coeso per poterlo definire una comunità?

5.1 Coefficiente di Clustering

Per misurare il grado di coesione di un nodo u è stato definito il **coefficiente di clustering** $c(u)$ come il rapporto fra il numero di relazioni esistenti fra i vicini di u e il numero di tutte le possibili relazioni tra di essi:

$$c(u) = \frac{|\{(x,y) \in E : x \in N(u) \wedge y \in N(u)\}|}{\frac{|N(u)|(|N(u)|-1)}{2}}$$

Informalmente, misura quanto un nodo è "ben inserito" nella rete dei suoi vicini.

Sia C un sottoinsieme di nodi. Definiamo il **coefficiente di clustering di u relativo a C** come:

$$c_C(u) = \frac{|\{(x,y) \in E : x \in N(u) \cap C \wedge y \in N(u) \cap C\}|}{\frac{|N(u) \cap C|(|N(u) \cap C|-1)}{2}}$$

Se tutti i nodi in C hanno un coefficiente di clustering relativo a C elevato, possiamo pensare che C sia una comunità.

6 Definizioni di Comunità

Esistono numerose definizioni di comunità. In queste lezioni ne vedremo alcune, tra cui:

- cut-communities;
- web-communities;
- metodi agglomerativi e partitivi per l'individuazione di comunità, basati su indici di centralità come la *betweenness*.

6.1 Cut-Communities

Definizione 4 (Cut-Community)**.** Una **cut-community** per un grafo $G = (V, E)$ è un sottoinsieme proprio e non vuoto C dei nodi di G che minimizza gli archi del taglio, ossia gli archi che collegano nodi in C a nodi in $V - C$. Formalmente, $C \subset V$ è una cut-community se $C \neq \emptyset$, $C \neq V$ e

$$|\{(u,v) \in E : u \in C \wedge v \in V - C\}| = \min_{C' \subset V, C' \neq \emptyset, C' \neq V} |\{(u,v) \in E : u \in C' \wedge v \in V - C'\}|$$

Calcolare un taglio minimo tra una coppia di nodi (s,t) è un problema computazionalmente trattabile (ad es. con l'algoritmo di Ford-Fulkerson). Per trovare una cut-community globale, si può iterare questo processo su tutte le coppie di nodi e scegliere il taglio con il valore minimo assoluto.

Problema delle Cut-Communities Tuttavia, questo approccio non permette di "controllare" le proprietà degli insiemi che costituiscono il taglio, in particolare la loro dimensione. Consideriamo il seguente esempio:

In questo grafo, i tagli minimi sono quelli che isolano un singolo nodo, tagliando 7 archi ciascuno. Quindi, una cut-community è un singolo nodo, il che non è una nozione di comunità molto ragionevole.

6.2 Web-Communities

Definizione 5 (Web-Community)**.** Una **(strong) web-community** è un sottoinsieme di nodi $C \subset V$ in cui ciascun nodo ha più vicini all'interno di C che all'esterno. Formalmente, $C \neq \emptyset$ e

$$\forall u \in C : |N(u) \cap C| > |N(u) \cap (V - C)| \quad \text{o, equivalentemente,} \quad \frac{|N(u) \cap C|}{|N(u)|} > \frac{1}{2}$$

Una **(weak) web-community** usa la disuguaglianza non stretta ($\geq$).

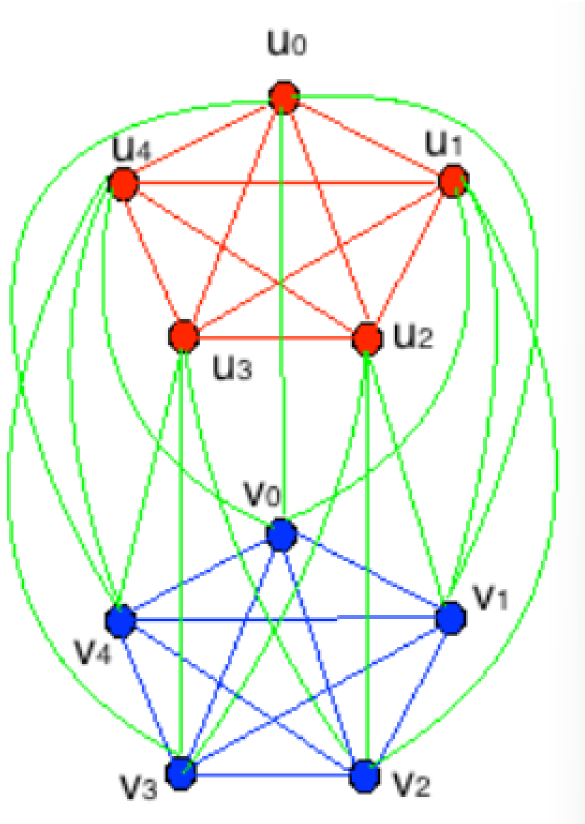

Figure 1: Un grafo G unione di due clique A (nodi U) e B (nodi V), con collegamenti tra di esse. Ogni nodo ha grado 7.

6.3 Relazione tra Cut- e Web-Communities

Le due definizioni non sono scorrelate.

Teorema 2. *Sia $G = (V, E)$ un grafo. Se $C \subset V$ è una cut-community tale che $|C| > 1$, allora C è una weak web-community.*

Proof. Supponiamo per assurdo che C non sia una weak web-community. Allora esiste un nodo $u \in C$ tale che $|N(u) \cap C| < |N(u) \cap (V - C)|$. Consideriamo il nuovo insieme $C' = C - \{u\}$. Poiché $|C| > 1$, C' è non vuoto. Calcoliamo la dimensione del taglio indotto da C':

$$\begin{aligned} |\text{taglio}(C')| &= |\{(x, y) \in E : x \in C \wedge y \in V - C\}| - |N(u) \cap (V - C)| + |N(u) \cap C| \\ &= |\text{taglio}(C)| - (|N(u) \cap (V - C)| - |N(u) \cap C|) \end{aligned}$$

Poiché per ipotesi $|N(u) \cap (V - C)| - |N(u) \cap C| > 0$, si ha $|\text{taglio}(C')| < |\text{taglio}(C)|$. Questo contraddice l'ipotesi che C sia una cut-community, poiché abbiamo trovato un taglio di dimensione inferiore. $\square$

Tuttavia, come visto nell'esempio precedente, gli algoritmi per le cut-communities possono restituire comunità di un solo nodo, e una cut-community di un solo nodo non è una web-community. Nel nostro esempio, $C = \{U_0\}$ è una cut-community ma non una web-community, poiché U_0 ha 0 vicini in C e 7 in $V - C$.

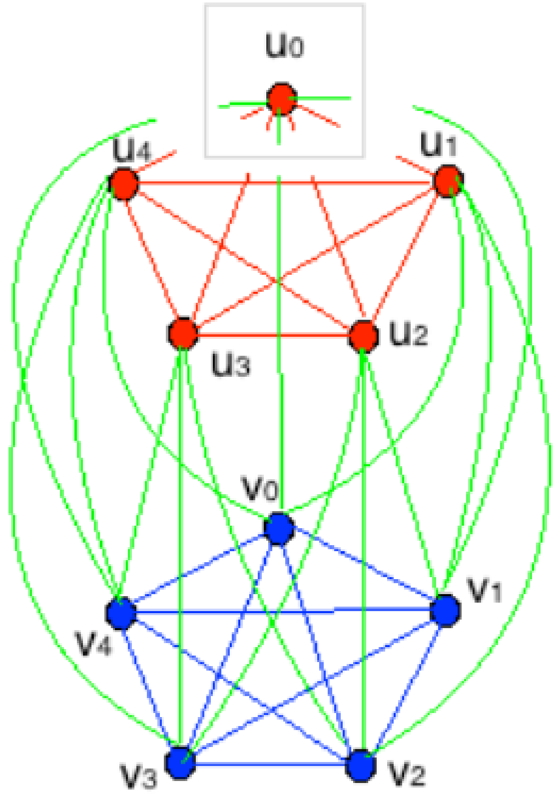

Figure 2: Il taglio minimo isola un singolo nodo, come U_0.

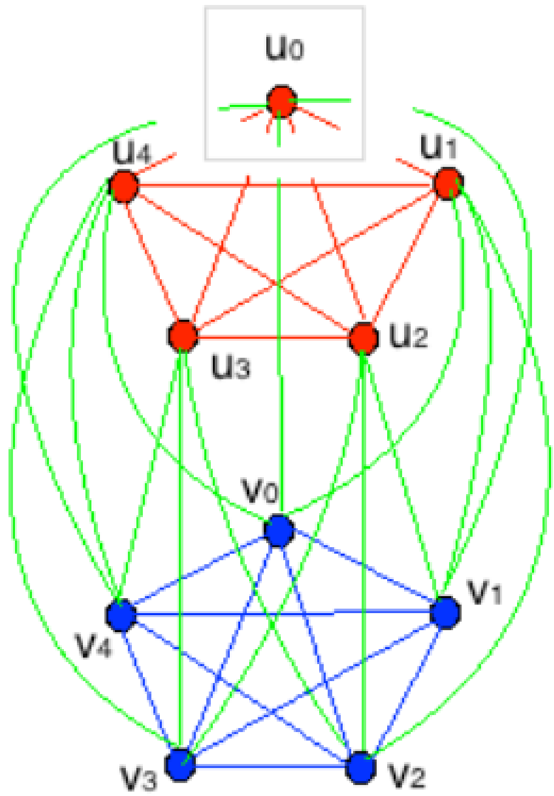

Figure 3: La cut-community $C = \{U_0\}$ non è una web-community.