

# Complex Network Tools for evaluation of P2P Networks

Matteo Sereno

Dipartimento di Informatica,  
Università di Torino

Alessandria  
4 Aprile, 2006



# Introduzione

- In un sistema P2P i peers posseggono (e offrono per la condivisione) risorse; la stessa risorsa può essere offerta da diversi peer
- Ogni risorsa è caratterizzata dalla sua popolarità
  - è una misura della diffusione di una risorsa
- Le attività tipiche di un peer sono:
  1. ricercare una risorsa (effettuando delle "query" ad altri peer)
  2. ottenere una copia della risorsa richiesta
  3. "processing" della risorsa
- I servizi offerti consistono nel partecipare alla fase di ricerca e fornire risorse (o porzioni di risorse) ai peer richiedenti

# Introduzione

- In un'applicazione P2P i peer si (auto)organizzano in una rete di overlay
- Ogni peer stabilisce connessioni verso un sotto-insieme di altri peer dell'applicazione (vicini)
- La gestione della rete di overlay viene effettuata al livello applicativo
- In questo lavoro ci occuperemo di applicazioni P2P di tipo decentralized unstructured
  - non esiste un gestore centralizzato del sistema (alla Napster)
- Per cercare una risorsa un peer deve inoltrare messaggi di richiesta (query) ai suoi vicini

**Proponiamo un framework analitico per analizzare strategie di ricerca in applicazioni P2P di tipo decentralized unstructured**

# Introduzione

- Uno degli approcci più usati per la ricerca in P2P unstructured decentralized è il flooding
- Nel flooding, quando un peer richiede una risorsa invia una query ad i suoi vicini
- Un peer che riceve un messaggio di query:
  - Inoltra la richiesta ad i suoi vicini
    - I vicini propagano il messaggio di query fino ad un numero massimo (predefinito) di hop (TTL)
  - Se il peer possiede la risorsa cercata invia un messaggio di query-hit
  - Scarta il messaggio se ha già ricevuto la stessa query

# Introduzione

- Le strategie di flooding devono essere progettate con attenzione:
  - Ottenere un'elevata probabilità di successo della query
  - Minimizzare il numero di messaggi
    - per ottenere una maggiore scalabilità del processo di ricerca
    - per ridurre il traffico dovuto alla ricerca

# Introduzione

- La ricerca in una rete P2P di tipo unstructured decentralized può essere visto come un esempio di processo dinamico su una rete complessa
- La ricerca su complex-network ha prodotto una serie numerosa di lavori/proposte/risultati che hanno evidenziato le relazioni tra proprietà topologiche della rete e caratteristiche del processo
- Diffusione di malattie, information dissemination, diffusione di virus mediante email, ecc.
- Una delle proprietà topologiche più importanti nelle reti complesse è la distribuzione del grado di un nodo

# Introduzione

- Distribuzione del grado in una rete di overlay di applicazioni P2P
- I primi lavori che hanno studiato (mediante misurazioni) distribuzione del grado della rete di overlay di Gnutella hanno mostrato che la distribuzione del grado è di tipo power-law (heavy-tailed)
- Altre misurazioni successive hanno evidenziato che in realtà la distribuzione del grado è di tipo bi-modale
- Alcuni studi recenti sulla rete Gnutella2 hanno evidenziato che la distribuzione del grado non è di tipo power-law
- Si evince la necessità di un framework matematico flessibile in grado di gestire differenti distribuzioni del grado di una rete di overlay

# I Generalized Random Graphs:

- Proposta:
  - Utilizzo dei Generalized Random Graphs (GRG) come formalismo matematico per modellare/rappresentare la rete di overlay network di applicazioni P2P di tipo unstructured decentralized
  - Uso del framework proposto per studiare strategie di ricerca flooding based
  - Estensioni varie
- Un GRG è composto da  $N$  nodi
  - distribuzione del grado di un nodo arbitraria
  - Il numero di archi uscenti da un nodo viene determinato dalla distribuzione del grado



# Assunzioni

- La probabilità di "trovare" due o più volte lo stesso nodo durante il processo di diffusione della query è trascurabile
  - Il numero di nodi incontrati durante il processo di diffusione della query viene rappresentato mediante variabili casuali indipendenti
- La scala temporale della diffusione di una query è piccola rispetto alla scala temporale tipica dell'evoluzione della rete di overlay

## Strategie di ricerca

- La ricerca basata su flooding viene analizzata come un processo di "spreading" di un contenuto (o di un virus)
- Strumenti analitici basati sull'uso di funzioni generatrici
- Valutazione di strategie di flooding
  - al variare di TTL
  - popolarità di risorse
  - strategie di flooding "maggiormente controllate"
- Validazione del modello

# Conclusioni

- Framework matematico per valutare strategie di ricerca flooding-based applicazioni P2P
- La proposta è derivata da approcci presentati in ambito di meccanica statistica
- L'analisi viene effettuata mediante l'uso di funzioni generatrici (efficiente dal punto di vista computazionale)
- Indici che possiamo ricavare sono
  - Numero di messaggi medio per query generata
  - Probabilità di trovare la risorsa (hit probability)
  - Distanza (in termini di hop) dalla prima risposta positiva
  - Tempi per ottenere risposte
- Il framework è in fase di estensione:
  - Generalized probabilistic flooding
  - Risorse distribuite in maniera non-uniform resource assignment
  - Comportamenti on-off dei peer

# Grazie per l'attenzione

## Riferimenti:

1. Gaeta, Balbo, Bruell, Gribaudo, Sereno, "A simple analytical framework to analyze search strategies in large-scale peer-to-peer networks", Performance Evaluation 2005
2. Gaeta, Sereno, "Model-based Evaluation of Search Strategies in peer-to-peer Networks", (Hot-P2P 2006)
3. Gaeta, Sereno, "An Analytical model to evaluate heterogeneous peer-to-peer networks, tech. report.



## Esempio di utilizzo del modello (II)

- Consideriamo:
  - Una distribuzione del grado di tipo power law with exponential cut-off
  - Due valori di popolazione  $p_k = \frac{k^{-\tau} e^{-k/\kappa}}{\text{Li}_\tau(e^{-1/\kappa})}$ ,  $\tau = 2.1$ ,  $\kappa = 500$
  - Due valori "target" di  $p_{\text{hit}}$  ( $p_{\text{min}}=0.5$  and  $p_{\text{min}}=0.75$ )
  - TTL che variano da 4 a 6
- Cosa si può osservare:
  - Le strategie di tipo  $g_2$  danno origine a risultati migliori di quelli che si ottengono usando strategie di tipo  $g_1$  (a parità di  $p_{\text{hit}}$  generano in media un numero minore di messaggi)
  - Altre considerazioni che si evincono al variare di parametri ( $\beta$  e  $p_{\text{min}}$ ) consentono di selezionare i set di parametri che consentono di ottenere le migliori prestazioni