



# Link Prediction su reti Multidimensionali

*Candidato*  
*Giulio Rossetti*

*Relatore*  
*Dott.ssa Fosca Giannotti*

*Correlatore*  
*Dott. Michele Berlingiero*

*Controrelatore*  
*Dott.ssa Chiara Bodei*

Pisa, 11 Marzo 2011



# Obiettivi e contributi della tesi

- Obiettivi
  - Definizione del problema di Link Prediction Multidimensionale
  - Formulazione di approcci algoritmici
- Contributi
  - Definizione ed Implementazione dei predittori
  - Costruzione delle reti multidimensionali
  - Estensiva analisi sperimentale
  - Rilascio plugin per Cytoscape





## Introduzione

Reti Multidimensionali e Multigrafi  
Definizione del problema

## Approcci di Link Prediction

Tipologie di approcci  
Modelli Derivati  
Modelli Ad Hoc

## Analisi Sperimentale

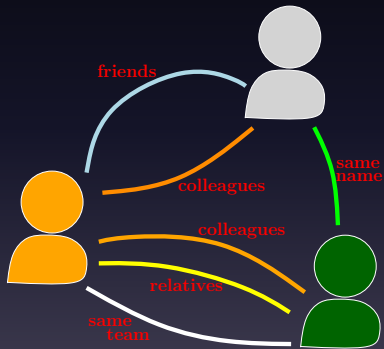
Reti multidimensionali analizzate  
Risultati  
Analisi Visuale

## Conclusioni

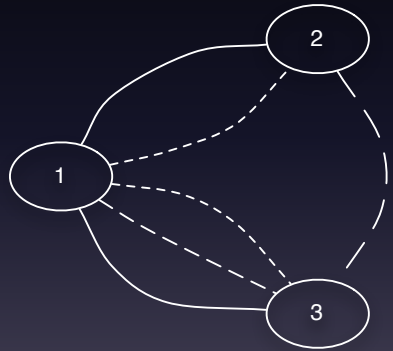




# Reti Multidimensionali e Multigrafi



Rete multidimensionale



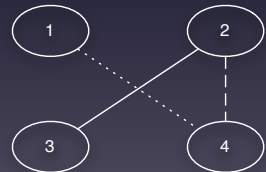
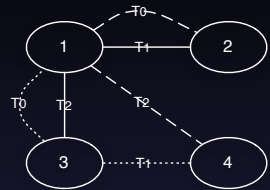
Multigrafo





# Link Prediction

- Dato
  - Multigrafo  $G = (V, E, D, T)$ 
    - $V$  insieme finito di vertici
    - $E$  insieme finito di archi
    - $D$  insieme finito di dimensioni
    - $T$  insieme finito di istanti temporali
- Trovare
  - Un modello predittivo che restituisca un insieme finito di quadruple  
(*nodo, nodo, dimensione, score*)





# Link Prediction: Approcci al problema

## 1 Approcci Unsupervised

- Basati su Neighborhoods
  - \* Common Neighbors, Adamic Adar, Jaccard...
- Basati su analisi dei Path
  - \* Shortest path, HITS, PR...
- ...

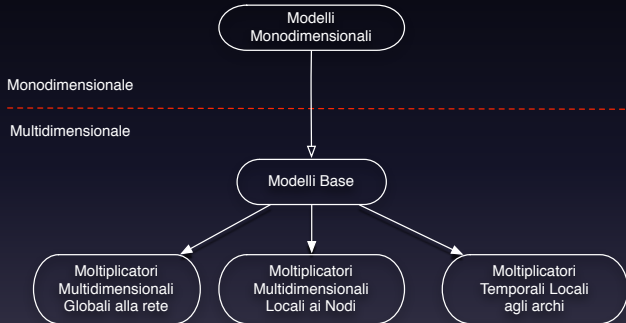
## 2 Approcci Supervised

- Guidati da task di Clustering/Classificazione
- Guidati dal modello evolutivo inferito dalla rete
  - GERM, LFR...
- ...

Approcci proposti nella tesi: unsupervised



# Approcci Proposti: Modelli Derivati



- 58 Modelli predittivi proposti
- Derivati da: Jaccard, Common Neighbors, Adamic Adar, Preferential Attachment

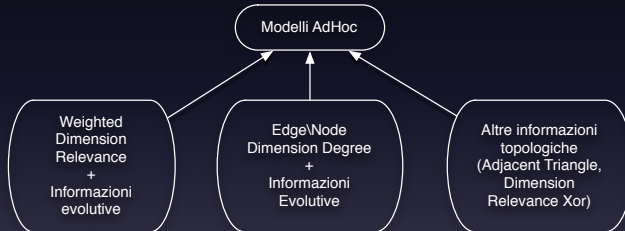




# Approcci Proposti: Modelli Ad Hoc

Monodimensionale

Multidimensionale



- 6 Modelli predittivi proposti
- Costruiti esclusivamente su misure multidimensionali e storia evolutiva della rete







## Reti utilizzate per l'analisi sperimentale - I

Rete	Nodi	Archi	Dimensioni
AOL Query Log	15.949	1.369.533	6
Internet Movie DB	43.867	216.544	28
DBLP	30.176	78.956	28





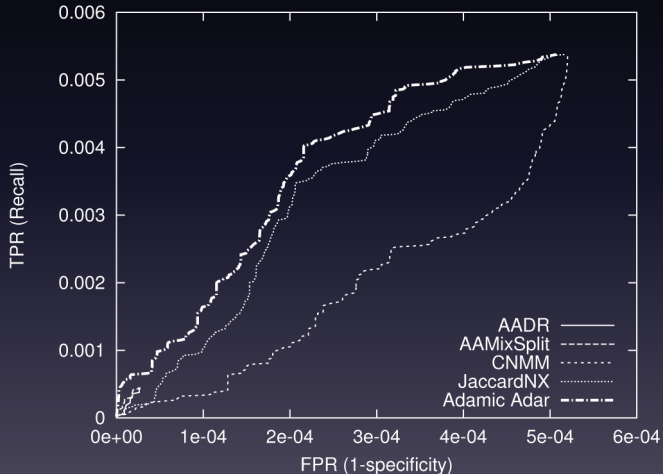
## Reti utilizzate per l'analisi sperimentale - II

Rete	Nodi	Archi	Dimensioni
Global Terrorism DB	2.755	25.200	209
Grand Comics DB	10.000	140.546	6
International Dyadic Events	320	413.766	247



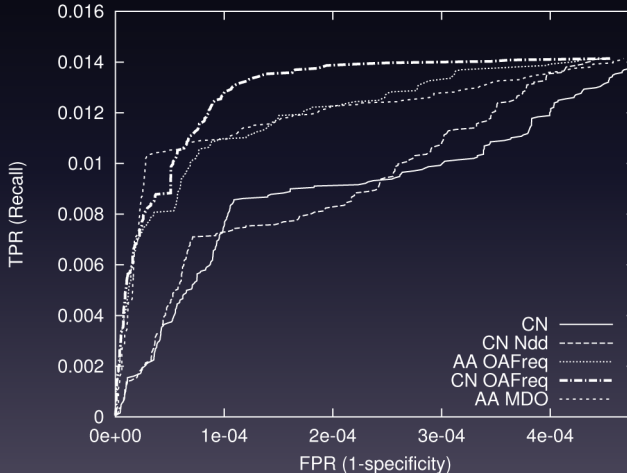


# Alcuni Risultati - I - Multidimensionali Locali



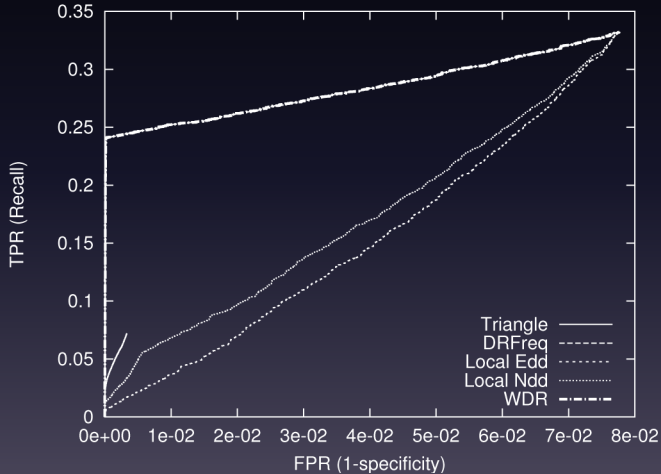


## Alcuni Risultati - II - Coeff. Globali e Temporal





## Alcuni Risultati - III - Ad Hoc



# Cytoscape Plugin

## Analisi visuale degli approcci predittivi proposti

The screenshot displays the Cytoscape Desktop (New Session) interface. The main window shows a network graph with five nodes (1-5) and several edges. A 'Multi Dimensional Link Prediction' dialog box is open, allowing the user to select a model for prediction. The 'Basic' tab is active, showing options for 'Local', 'Global', 'Evolutive', and 'Ad-hoc' models. Under 'Predictors (Neighbours for Version)', the 'Normal Version' is selected, with 'AA', 'CN', 'JC', and 'PA' as sub-options. The 'Compute' button is visible at the bottom of the dialog. The left sidebar contains the 'Control Panel' with instructions for adding nodes and edges, and a 'Data Panel' at the bottom.





## Cosa abbiamo imparato..

- Inesistenza di un predittore universale
- Modelli Derivati
  - Insufficienza informazioni multidimensionali locali ai nodi
  - Utilità informazioni multidimensionali globali alla rete
  - Aumenti di performance tramite analisi della storia evolutiva
- Modelli Ad Hoc
  - Migliori performance predittive ottenute

