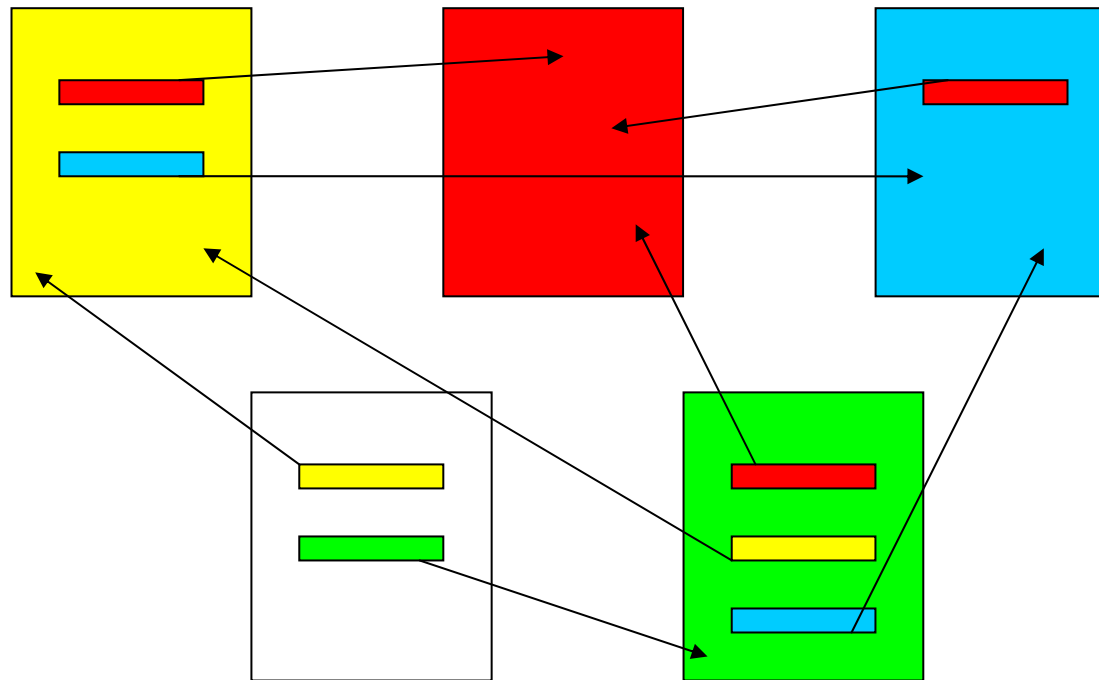
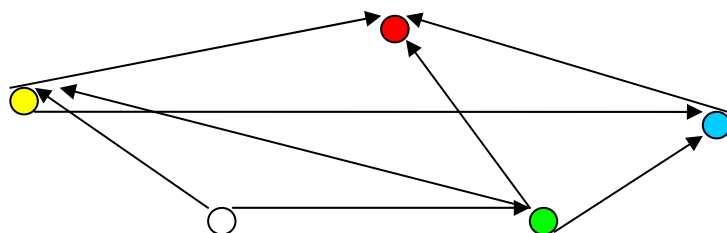


UNO STRUMENTO MATEMATICO: IL GRAFO

In corrispondenza dei link esistenti, disegniamo una freccia (**arco**) che parte dalla pagina “sorgente” e arriva alla pagina “destinazione”.



A questo punto non c'è più bisogno di disegnare esplicitamente i link dentro le pagine, e possiamo rappresentare ogni pagina con un punto colorato (**nodo**). L'insieme dei nodi e degli archi che li collegano si dice **grafo**, ed è uno strumento matematico di rappresentazione molto comodo (nonché molto generale).



Definizione. In un grafo, il *grado* di un nodo è il numero degli archi da esso uscenti.

nodo	●	●	●	○	●
grado	2	0	1	2	3

LA PASSEGGIATA CASUALE

Possiamo pensare di approssimare l'importanza di una pagina con la probabilità che la stessa ha di essere visitata da un navigatore virtuale che opera nel seguente modo:

- Ad ogni istante (minuto) il navigatore cambia pagina
- Per spostarsi da una pagina all'altra il navigatore decide di cliccare su uno dei link presenti sulla pagina corrente, scegliendolo a caso con uguale probabilità.

L'IDEA DI GOOGLE

- Ogni pagina ha una sua propria importanza che deriva dalle connessioni (non direttamente dai contenuti)
- L'importanza di una pagina viene trasferita in parti uguali alle pagine che essa punta
- L'importanza di una pagina è data dalla somma delle frazioni di importanza che gli derivano dalle pagine che ad essa puntano

UN ESEMPIO "IRRIVERENTE" ...

- Se il Papa nell'incontro della domenica a Piazza San Pietro dà la sua benedizione al professor Antonello Zibibbo, il professore riceve grande importanza
- Se il Papa dà la sua benedizione a tutti i professori del mondo allora il prof. Antonello Zibibbo riceve un'importanza trascurabile
- Se il professor Antonello Zibibbo benedice il Papa, quest'ultimo non se ne accorge nemmeno!

IL “TELETRASPORTO”

Il valore x_j del Page Rank calcolato da Google coincide con la probabilità che la pagina j ha di essere visitata da un navigatore virtuale che opera nel seguente modo:

- Ad ogni istante (minuto) il navigatore cambia pagina
- Per spostarsi da una pagina all'altra il navigatore:
 - con probabilità c decide di cliccare su uno dei link presenti sulla pagina corrente, scegliendolo a caso con uguale probabilità;
 - con probabilità $1 - c$ decide di saltare ad un'altra pagina a caso presente sul web, scegliendola con uguale probabilità.

In questo modo si garantisce che ogni pagina web può essere visitata da qualunque punto con probabilità non nulla.

...MA QUANTO TEMPO CI VUOLE?

Alla data di oggi ci sono circa $n = 8.5 \times 10^9$ pagine attive.

Se cerchiamo di risolvere il sistema di equazioni, il metodo di sostituzione (detto anche di Eliminazione Gaussiana) richiede poco più di $\frac{2}{3} n^3$ operazioni aritmetiche; nel nostro caso servirebbero circa

$$\frac{2}{3} \times (8.5 \times 10^9)^3 \approx 4.1 \times 10^{29}$$

(410 miliardi di miliardi di miliardi) operazioni aritmetiche.
Sono tante???

Anche disponendo del *Blue Gene* dell'IBM, per eseguire 4.1×10^{29} operazioni ci vorrebbero “solo” 36 milioni di anni!! Ma Google deve calcolare il page rank ogni mese...

...quindi NON risolve il sistema di equazioni; ne approssima la soluzione con le passeggiate casuali!

Per fare un passo della passeggiata casuale, bisogna eseguire tante moltiplicazioni quanti sono i link totali esistenti sul web, e all'incirca altrettante addizioni. Per fortuna, i link sono “pochi” in rapporto al caso peggiore in cui tutte le pagine si puntassero tra di loro (n^2).

Se mediamente ci fossero 50 link su ogni pagina, un passo del metodo iterativo eseguito col *Blue Gene* impiegherebbe 2 millesimi di secondo. Se anche fossero necessari 1000 passi casuali basterebbero 2 secondi per approssimare la soluzione di Google.