Padroni del Vapore

Gli algoritmi
(e chi li conosce)
dominano il mondo

Alberto Montresor

Trento, Ottobre 2018

Due parole su di me!

- Professore presso il Dipartimento di Ing. e Scienza dell'Informazione (Università di Trento)
- Titolare di tre corsi:
  - **Algoritmi e Strutture Dati** (LT Informatica)
  - **Scientific Programming** (LM Quantitative and Computational Biology e LM Data Science)
  - **Distributed systems 2** (LM Computer Science)

Discussione

Due parole su di voi!

Padroni del Vapore

Gli algoritmi
(e chi li conosce)
dominano il mondo

Alberto Montresor

Trento, Ottobre 2018

Padrone del vapore

L’origine è marinaresca: la parola ***padrone*** identifica il grado di chi è abilitato al comando di mercantili di una certa stazza

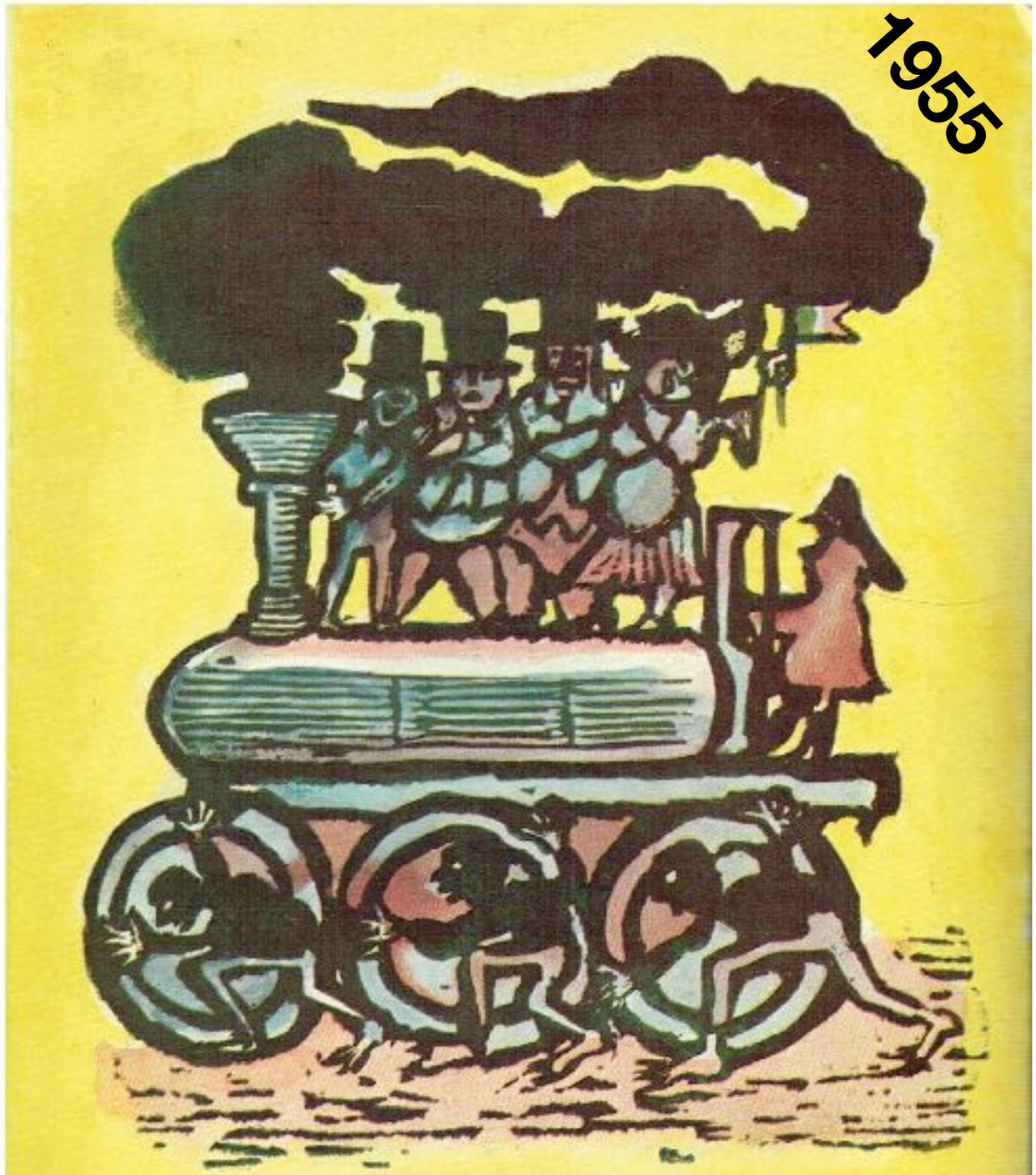
Dizionario dei modi di dire, Corriere.it:

(Fig.) Chi detiene il potere decisionale, quindi comanda, gestisce le cose a proprio giudizio e le impone agli altri.

Dizionario Il Nuovo De Mauro:

(Spreg.) Chi detiene il potere assoluto di un’organizzazione ed è *in grado di condizionare poteri pubblici*

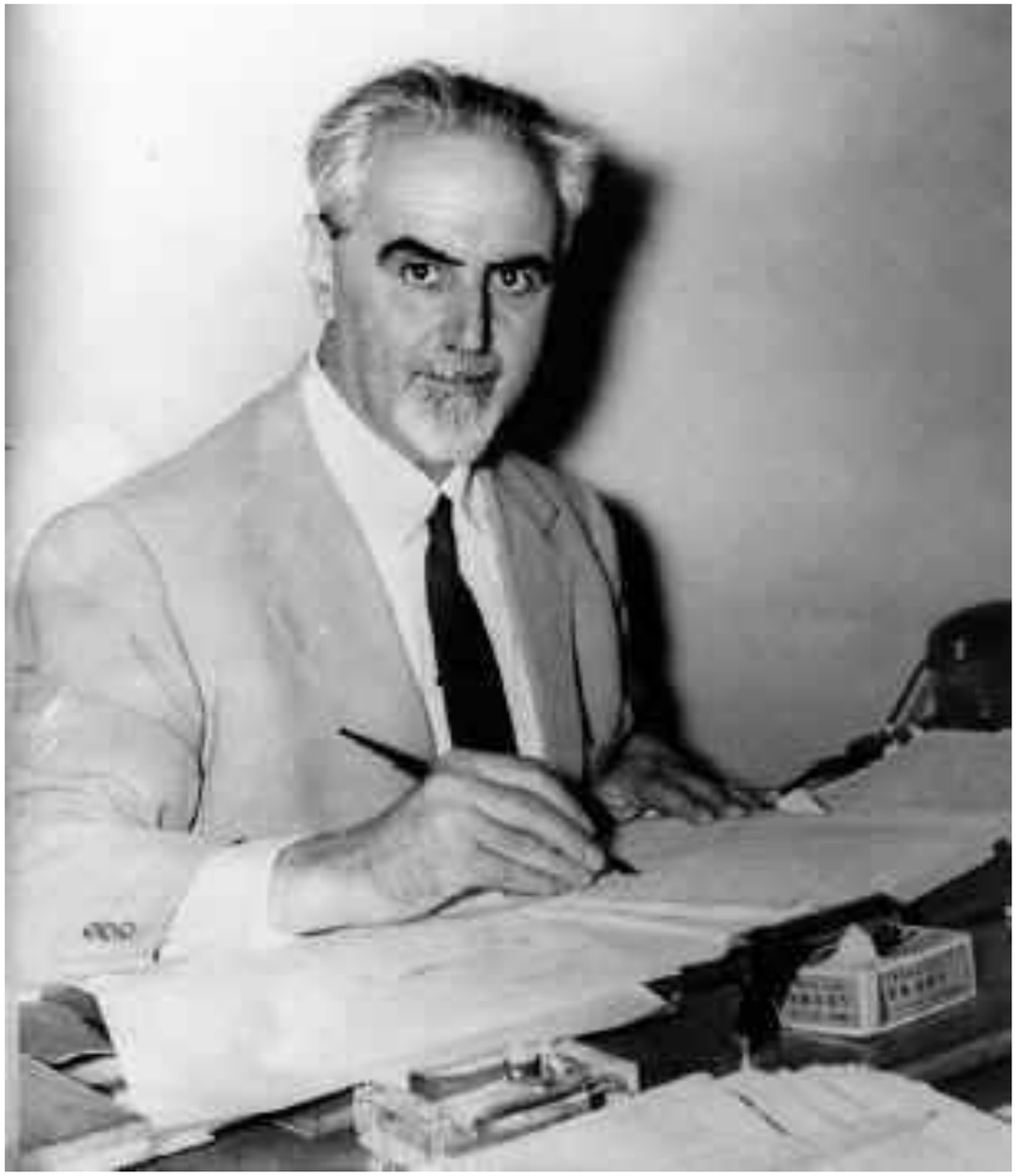
1955



Ernesto Rossi

I padroni del vapore

Editori Laterza




Person of the Year
TIME
Facebook's
Mark
Zuckerberg


TIME
COMPUTER SOFTWARE
The Magic Inside the Machine
Microsoft
Boss
Bill Gates


THE
100
MOST
INFLUENTIAL
PEOPLE
JEFF
BEZOS
by Buzz Aldrin


THE
100
MOST
INFLUENTIAL
PEOPLE
IN THE WORLD

Gli algoritmi dominano il mondo

La nostra vita è condizionata da centinaia di algoritmi e dalle decisioni che essi prendono per conto nostro.

THE WALL STREET JOURNAL

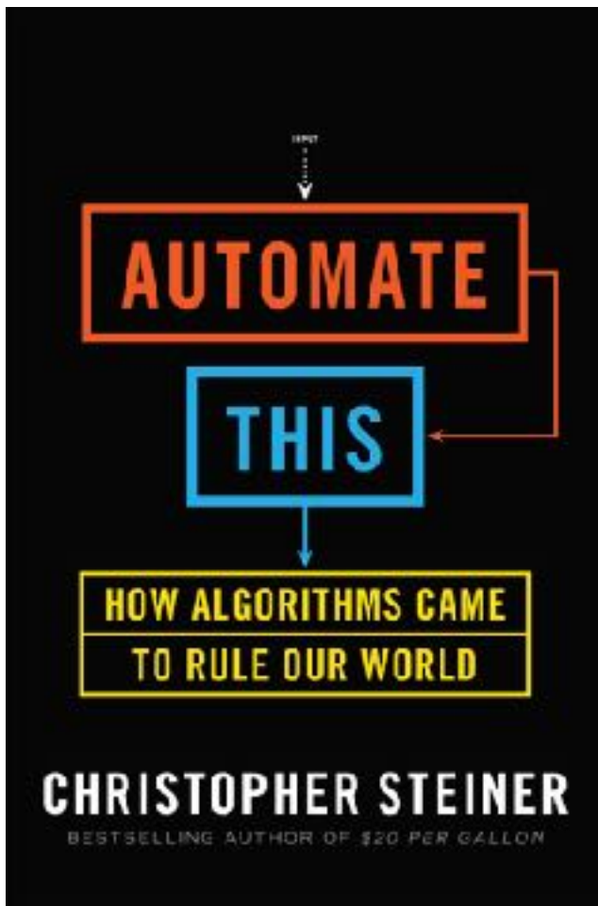
ESSAY

Why Software Is Eating The World

By Marc Andreessen

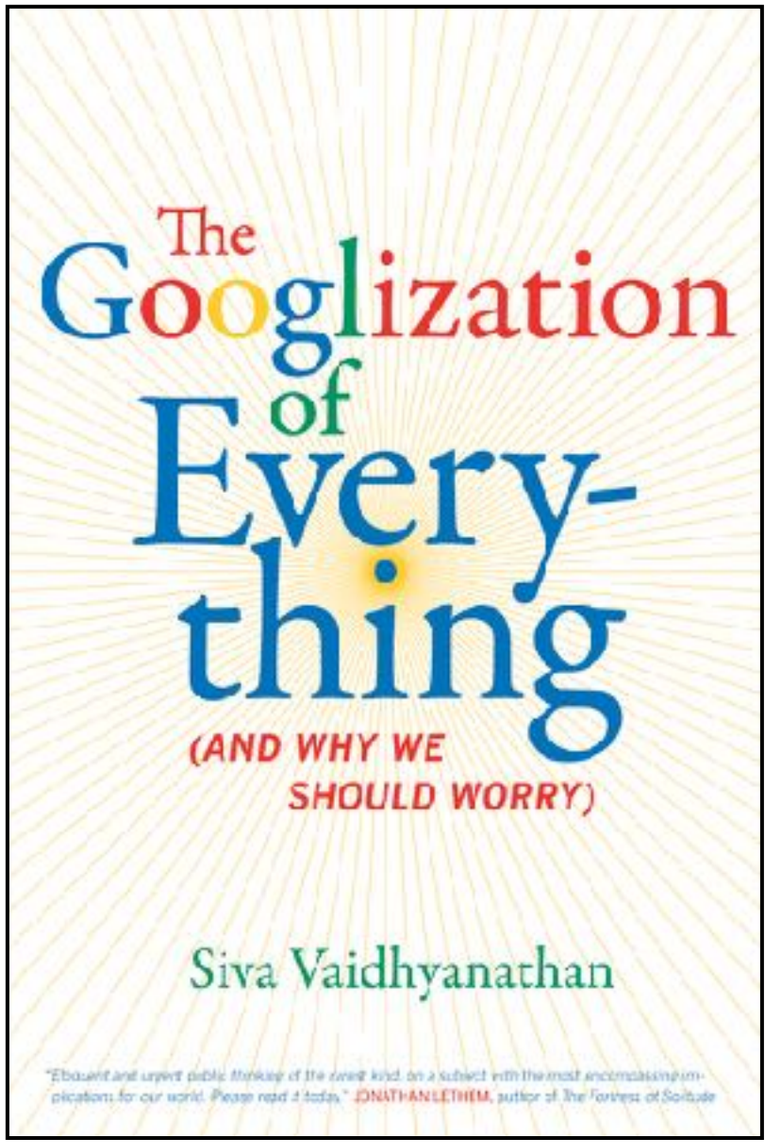
August 20, 2011

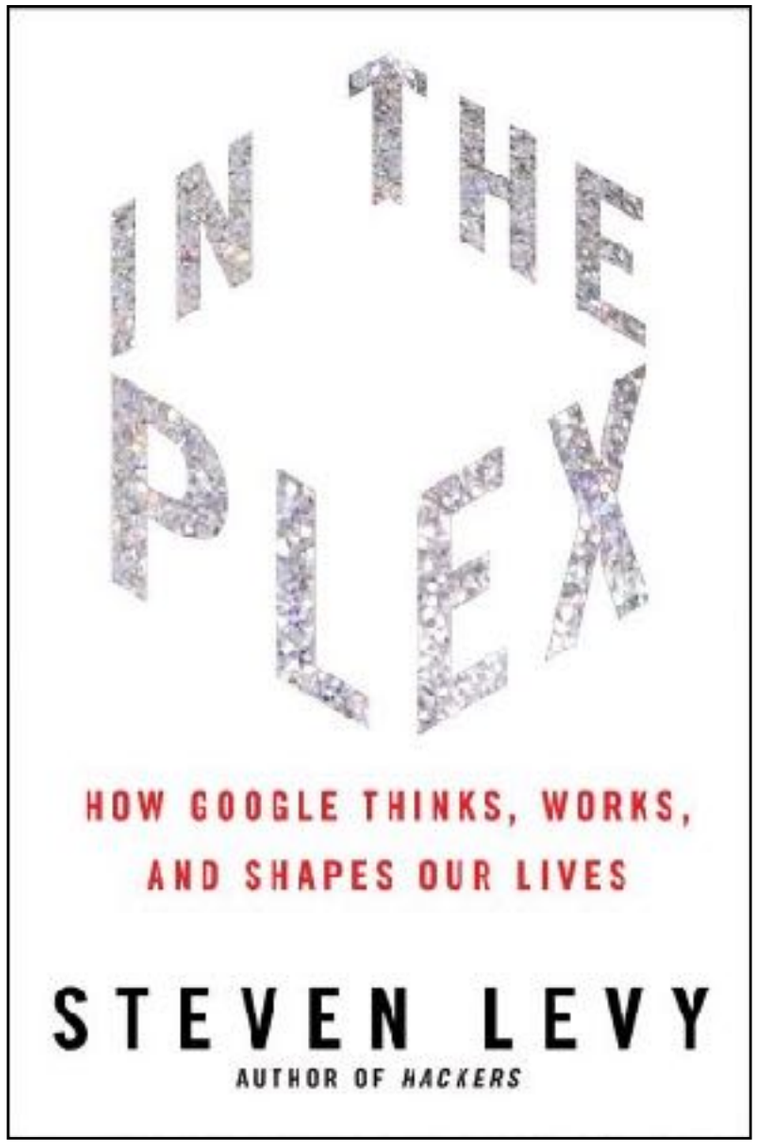
This week, Hewlett-Packard (where I am on the board) announced that it is exploring jettisoning its struggling PC business in favor of investing more heavily in software, where it sees better potential for growth. Meanwhile, Google plans to buy up the cellphone handset maker Motorola Mobility. Both moves surprised the tech world. But both moves are also in line with a trend I've observed, one that makes me optimistic about the future growth...



Chi controlla gli algoritmi ha un potere enorme

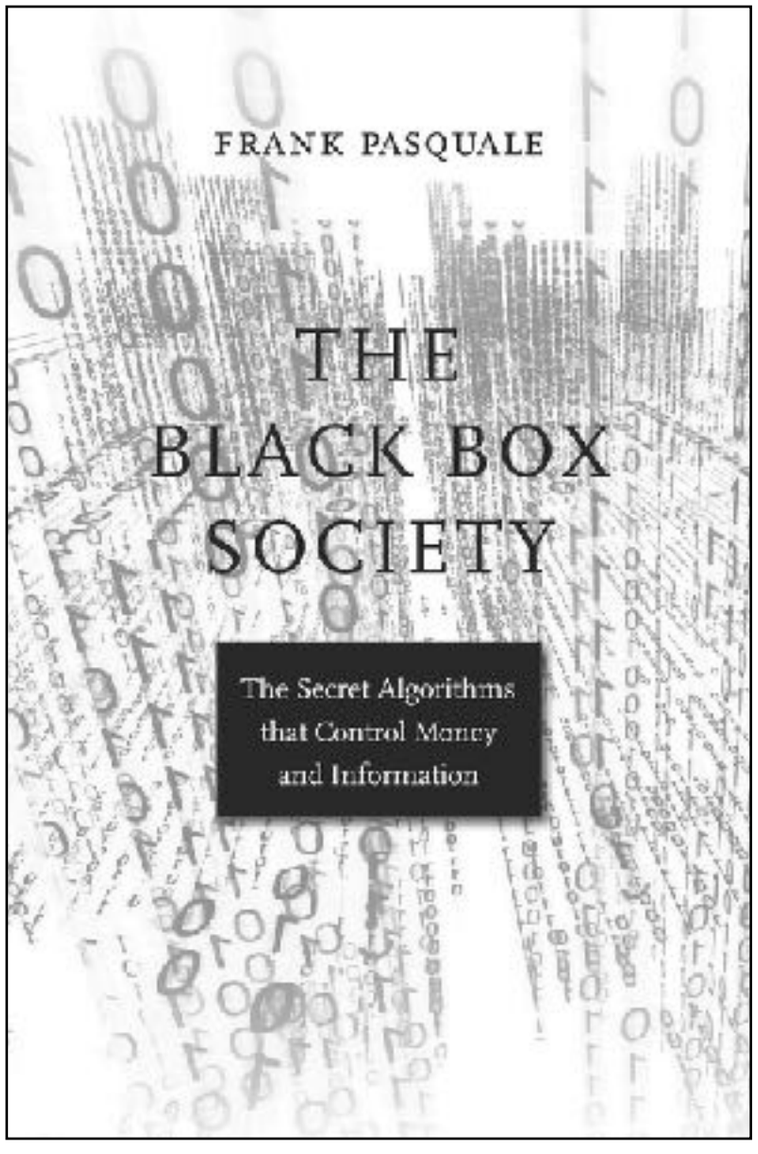
*Il potere decisionale si concentra sempre di più nelle mani di pochi "padroni del vapore", che hanno potere assoluto sulle **informazioni**, sui **prodotti**, in ultima analisi sulle **idee** che possono circolare.*

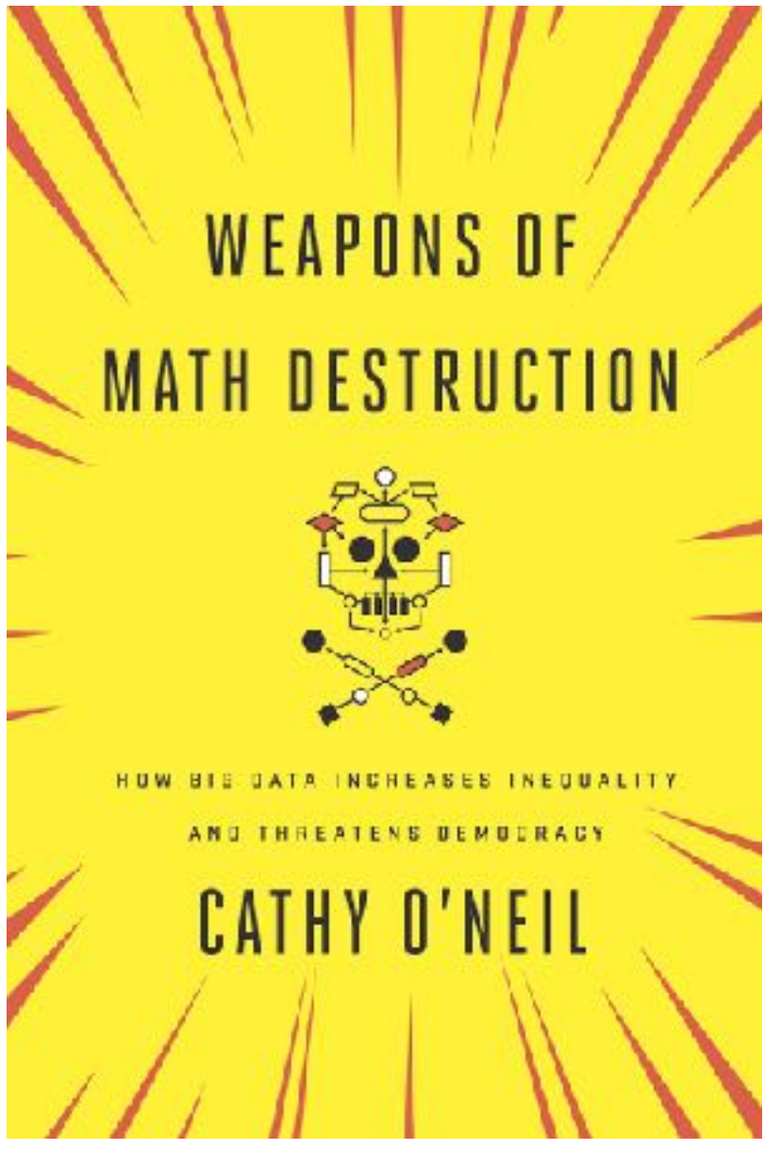




Non c’è trasparenza su come gli algoritmi esercitino il loro potere

*I motori di ricerca e i social network selezionano dai loro enormi database ciò che **essi** ritengono **più rilevante** per noi*





Esercitare il controllo sugli algoritmi diventa ogni giorno più difficile

*Gli algoritmi non vivono in isolamento. Sono parte di sistemi **socio-tecnici** complessi, che includono gli sviluppatori, gli esecutori, i dati su cui operano, gli algoritmi stessi, e infine gli utilizzatori.*

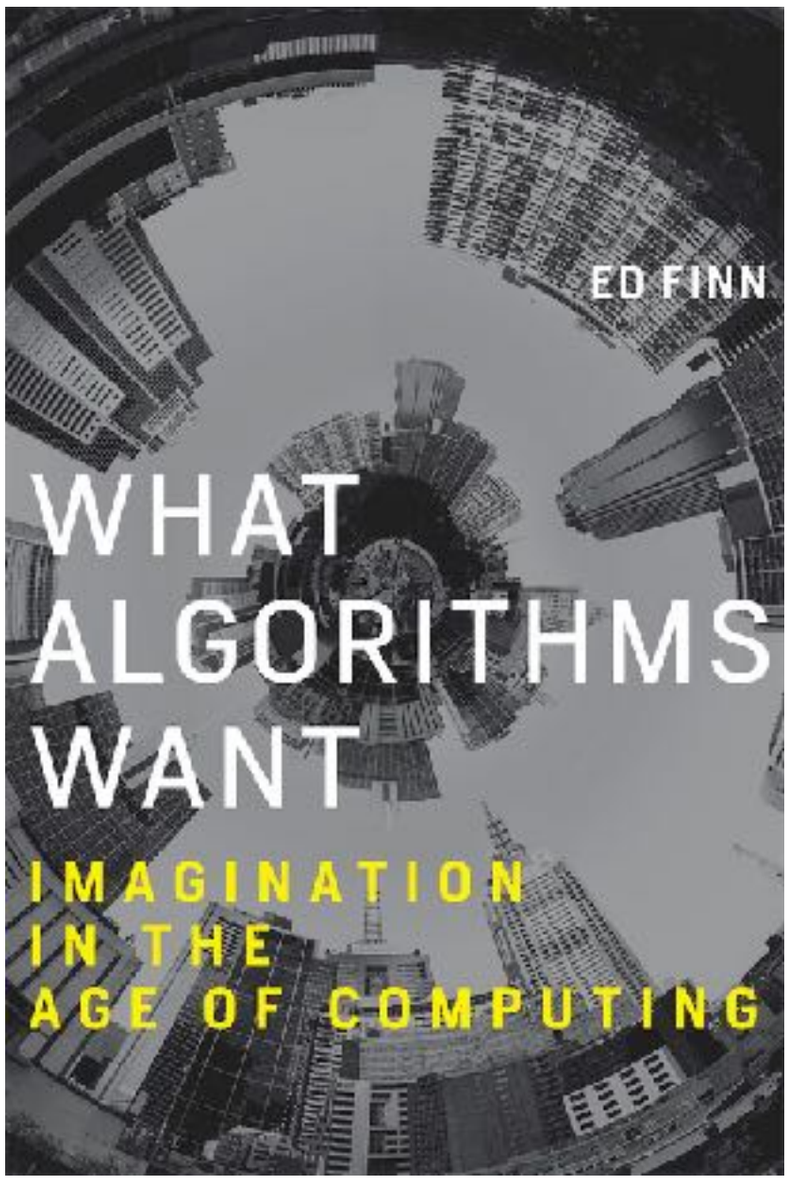
Forbes / Tech
JAN 4, 2016 @ 10:18 AM 8,241

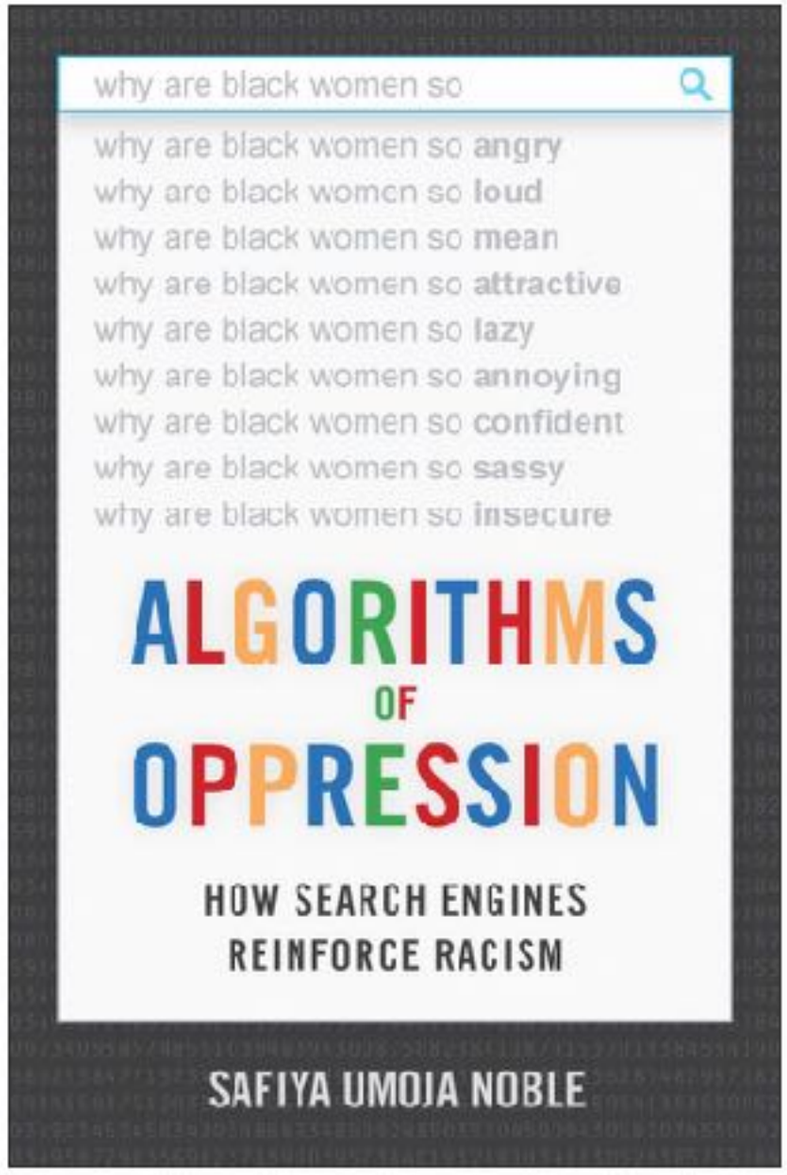
In Machines We Trust: Algorithms Are Getting Too Complex To Understand

Kalev Leetaru, CONTRIBUTOR
I write about the broad intersection of data and society. FULL BIO

Gli algoritmi hanno una “volontà propria”

Gli algoritmi riflettono in parte le volontà dei loro creatori, ma spesso assumono una volontà propria, che deriva dai loro utilizzatori.





The Web, circa 1991

The WWW Virtual Library : en · es · fr · zh

The WWW Virtual Library

Quick search:

Agriculture
Irrigation, Livestock, Poultry Science, …

The Arts
Art History, Classical Music, Theatre and Drama, …

Business and Economics
Finance, Marketing, Transportation, …

Communications and Media
Broadcasters, Publishers, Telecommunications, …

Computing and Computer Science
Artificial Intelligence, Cryptography, Logic Programming, …

Education
Primary, Secondary, Tertiary, …

Engineering
Architecture, Electrical, Mechanical, …

Humanities and Humanistic Studies
History, Languages and Linguistics, Museums, …

Information and Libraries
Information Quality, Knowledge Management, Libraries, …

International Affairs
International Relations and Security, Sustainable Development, …

Law
Arbitration, Forensic Toxicology, Legal History, …

Natural Sciences and Mathematics
Biosciences, Earth Science, Medicine and Health, Physics, …

Recreation
Gardening, Recreation and Games, Sport, …

Regional Studies
African, Asian, Latin American, European, …

Social and Behavioural Sciences
Anthropology, Archaeology, Population and Development Studies, …

Society
Peoples, Religion, Gender Studies, …

About | Contact | Donors | Topics

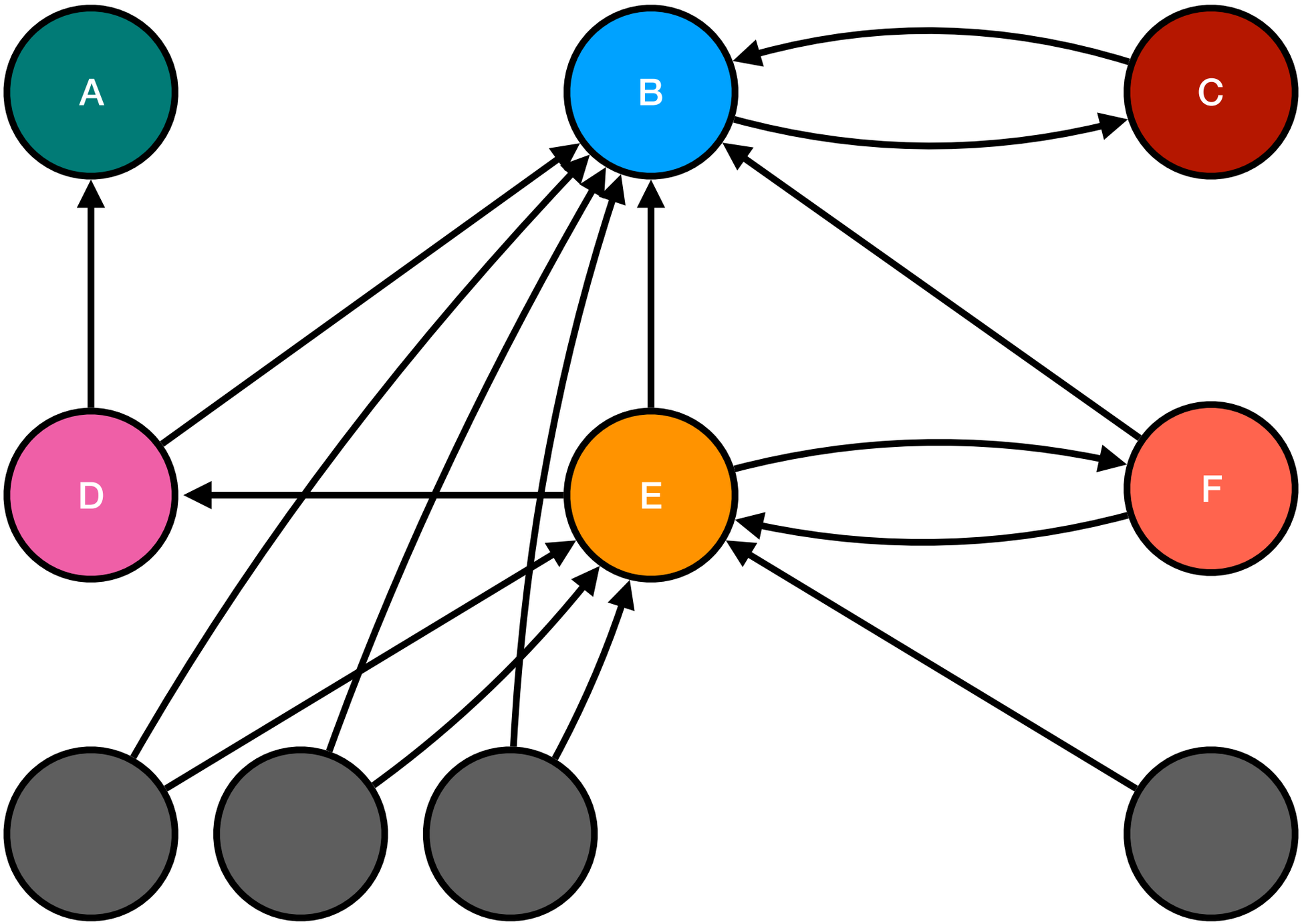
The Web, circa 1995

PageRank

Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd.
"*The PageRank citation ranking: Bringing order to the Web”.*
Technical Report, Stanford University, 29 January 1998.

Gli autori principali (Page e Brin)
propongono un nuovo meccanismo
per calcolare la *rilevanza* di una pagina
a partire dai link ad altre pagine

PageRank



Random Surfer Model

Dall’articolo originale:

*PageRank can be thought of as a model of user behavior. We assume there is a "**random surfer**" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. **The probability that the random surfer visits a page is its PageRank.***

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

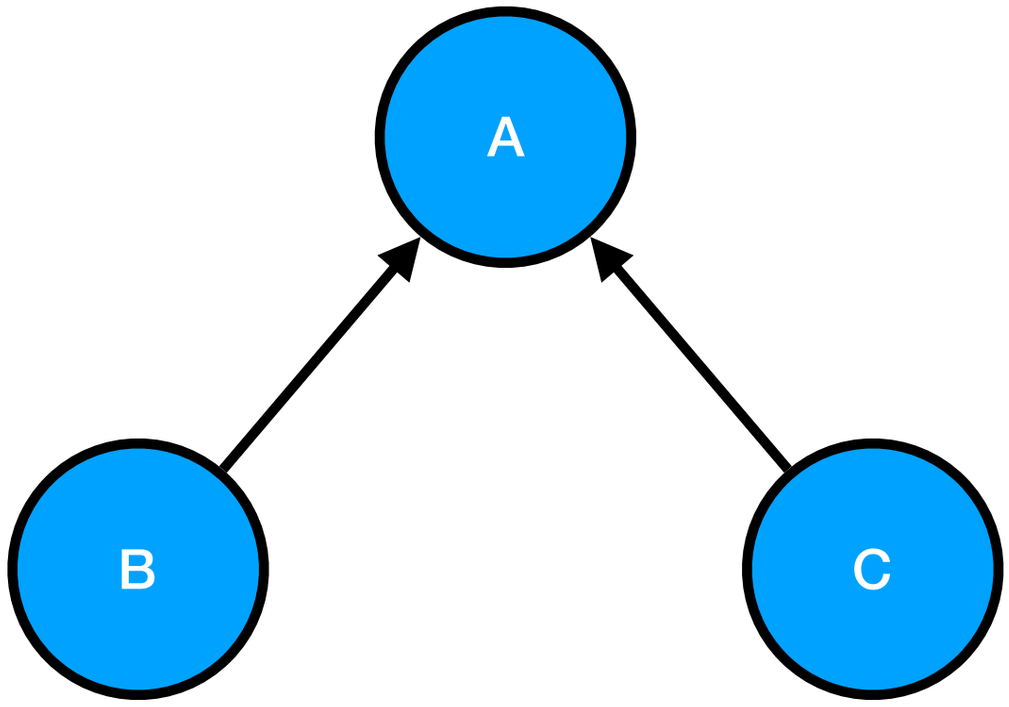
PageRank - Formula

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

- $PR(p_i)$ è la *rilevanza* della pagina *i*-esima
- N è il numero totale di pagine
- $M(p_i)$ è l'insieme di pagine che hanno un link verso p_i
- $L(p_j)$ è l'out-degree di p_j, ovvero il numero totale di link della pagina p_j
- *d* è detto damping factor, pari al valore empirico 0.85 (85%)

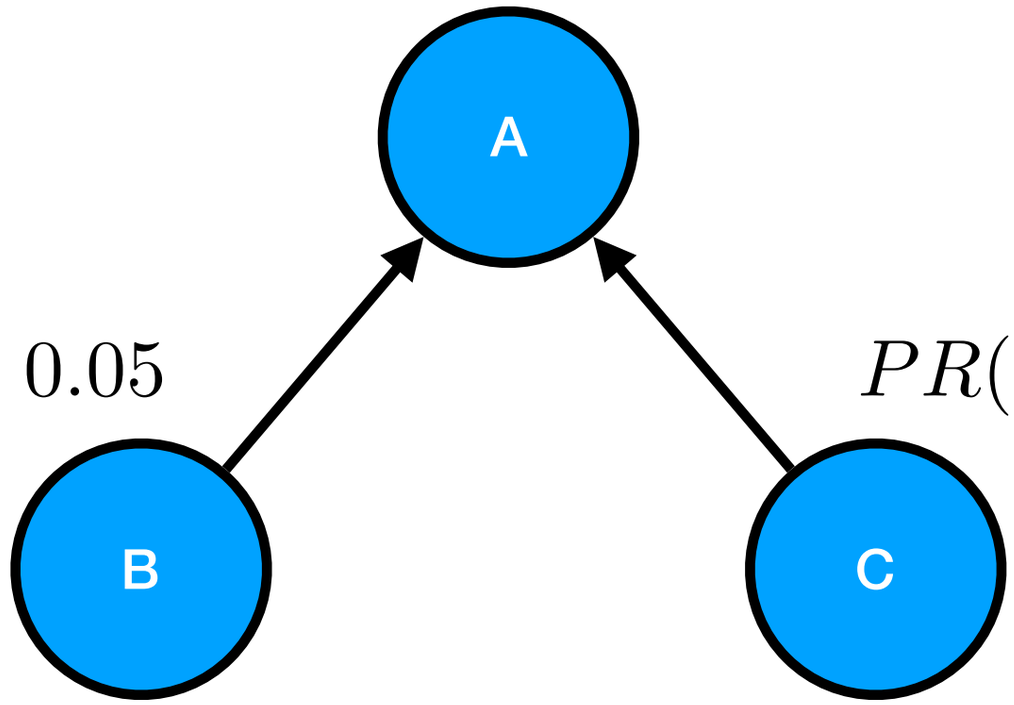
PageRank - Casi base

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$



PageRank - Casi base

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$



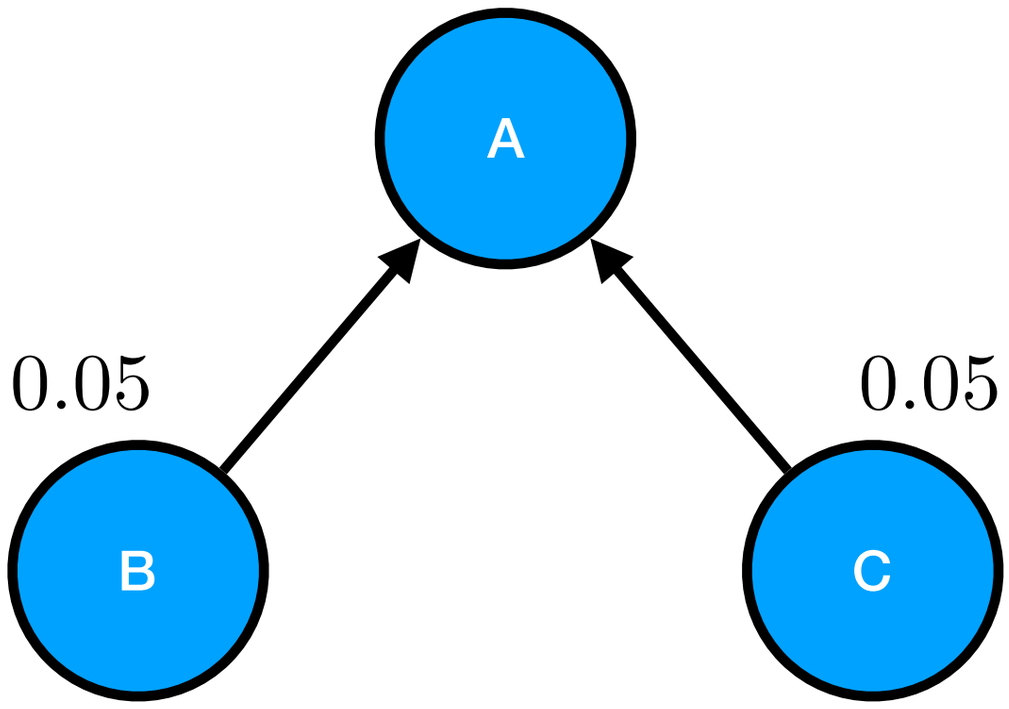
$$PR(B) = \frac{1-0.85}{3} = 0.05$$

$$PR(C) = \frac{1-0.85}{3} = 0.05$$

PageRank - Ricorsione

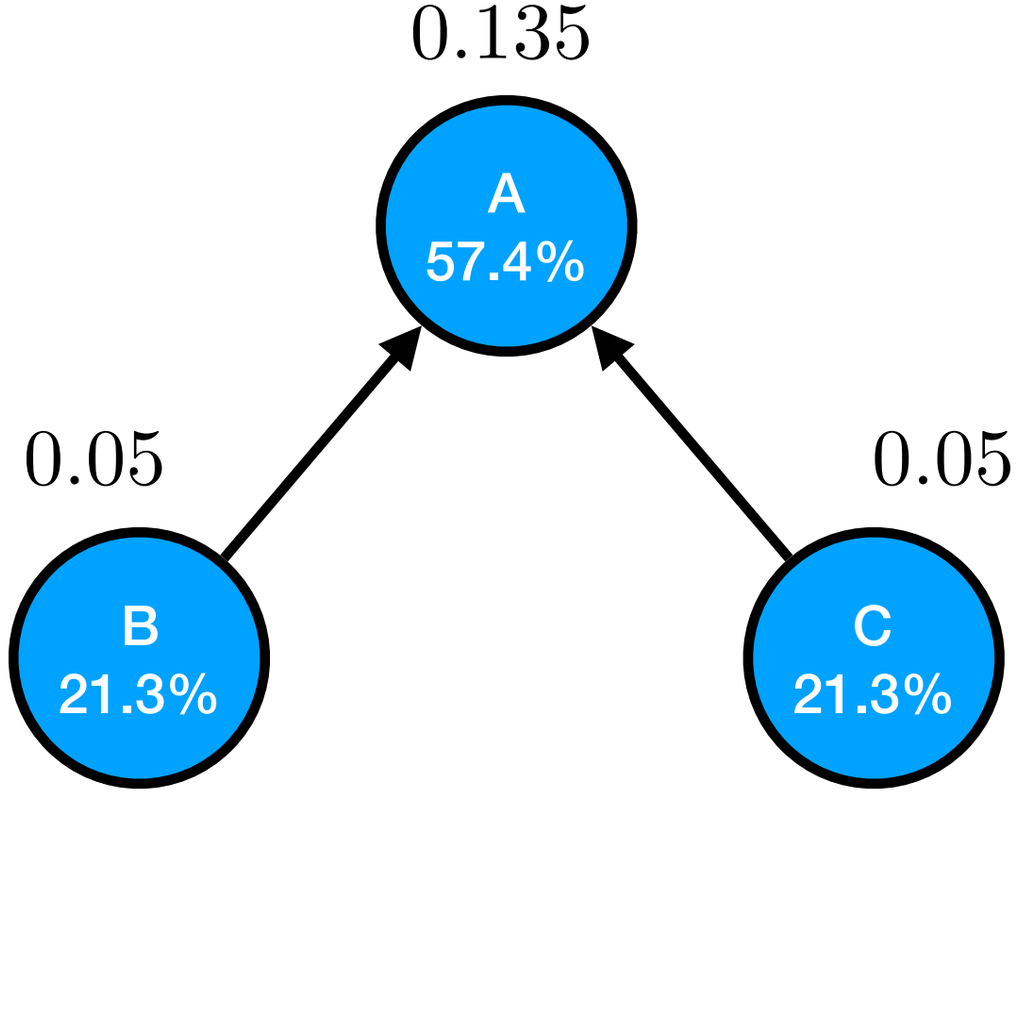
$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

$$PR(A) = \frac{1-0.85}{3} + 0.85 \left(\frac{PR(B)}{1} + \frac{PR(C)}{1} \right) = 0.135$$



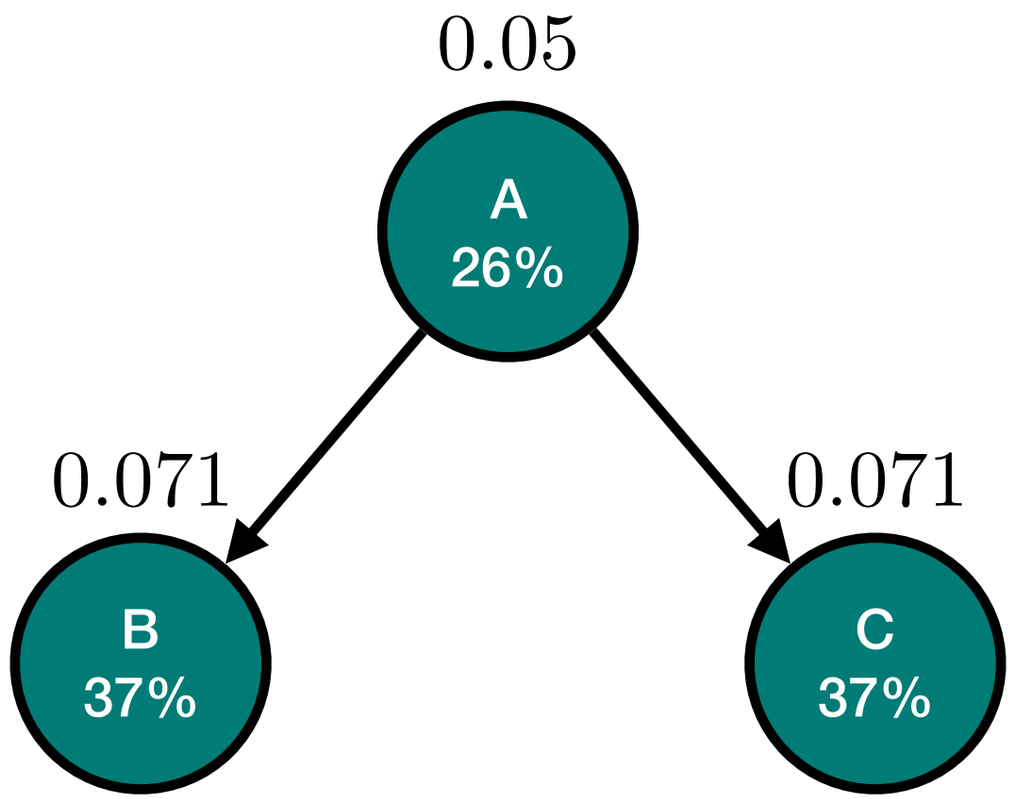
PageRank - Distribuzione

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$



PageRank - Casi base

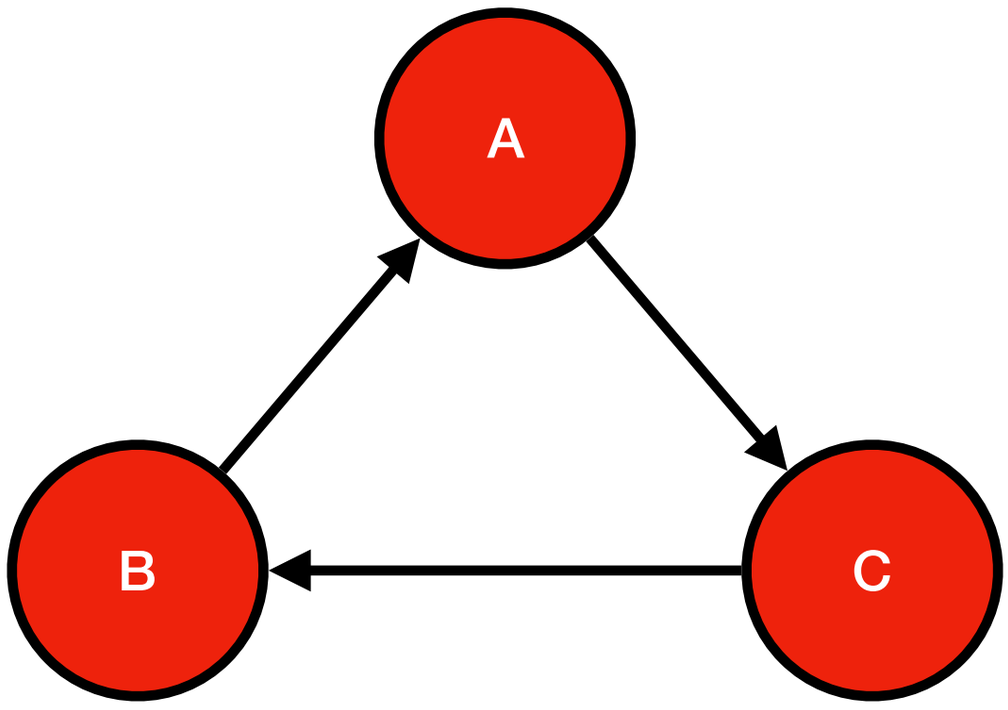
$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$



$$PR(B, C) = 0.05 + 0.85 \cdot \frac{PR(A)}{2}$$

PageRank - Cicli

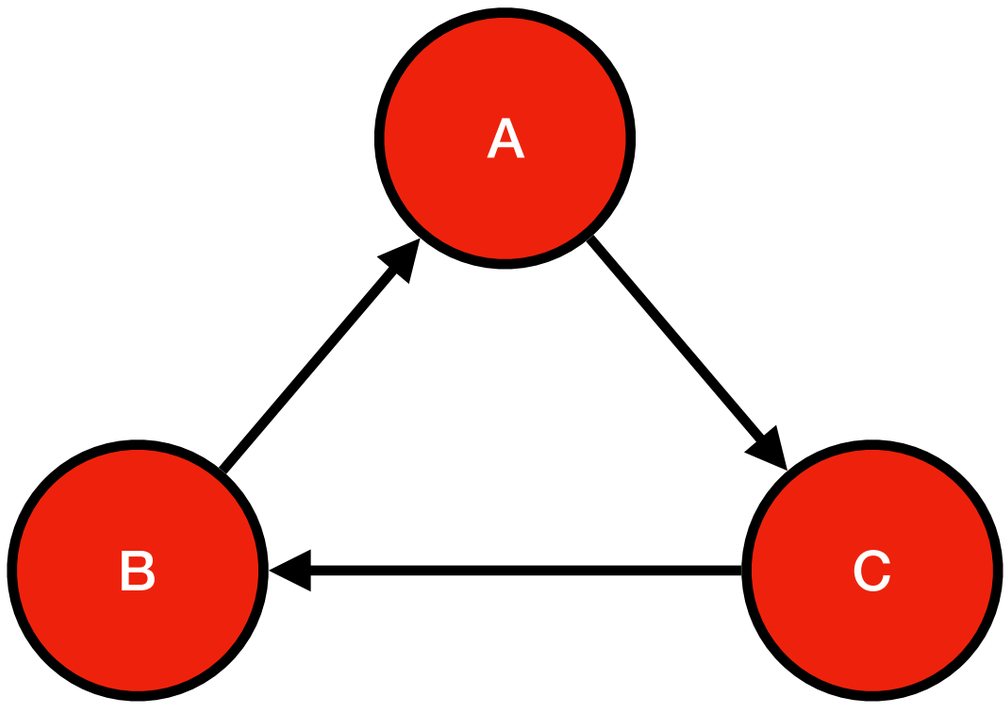
$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$



PageRank - Power iteration

$$PR(p_i, 0) = 1$$

$$PR(p_i, t) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j, t-1)}{L(p_j)}$$



PageRank - Power iteration

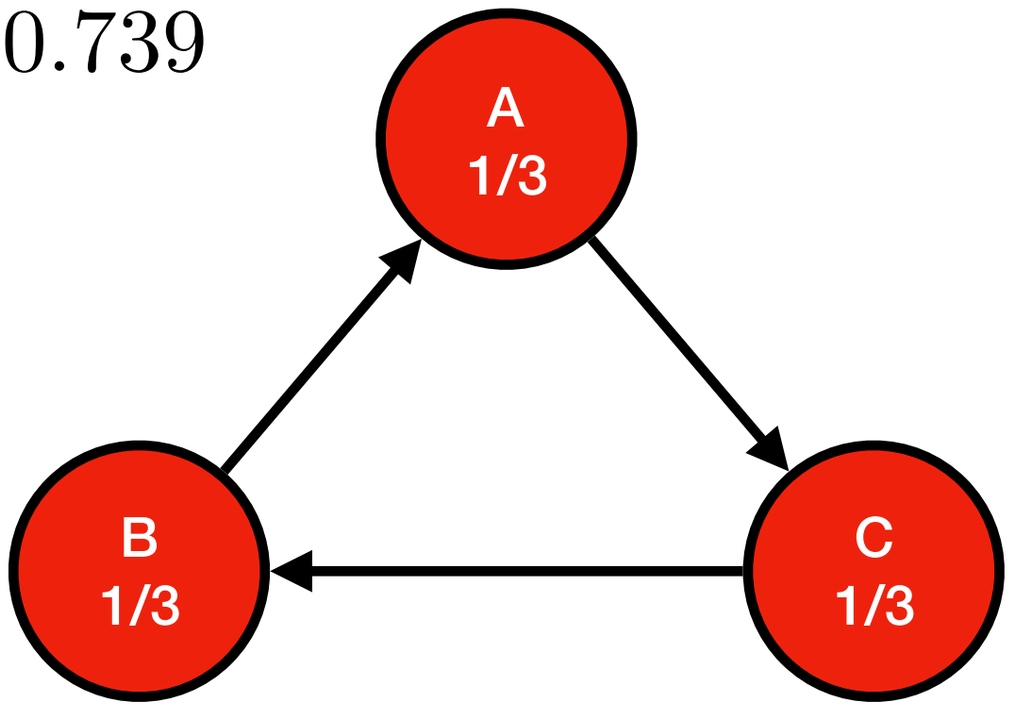
$$PR(x,0) = 1.000$$

$$PR(x,1) = \frac{1-0.85}{3} + 0.85 \cdot 1.000 = 0.900$$

$$PR(x,2) = \frac{1-0.85}{3} + 0.85 \cdot 0.900 = 0.810$$

$$PR(x,3) = \frac{1-0.85}{3} + 0.85 \cdot 0.810 = 0.739$$

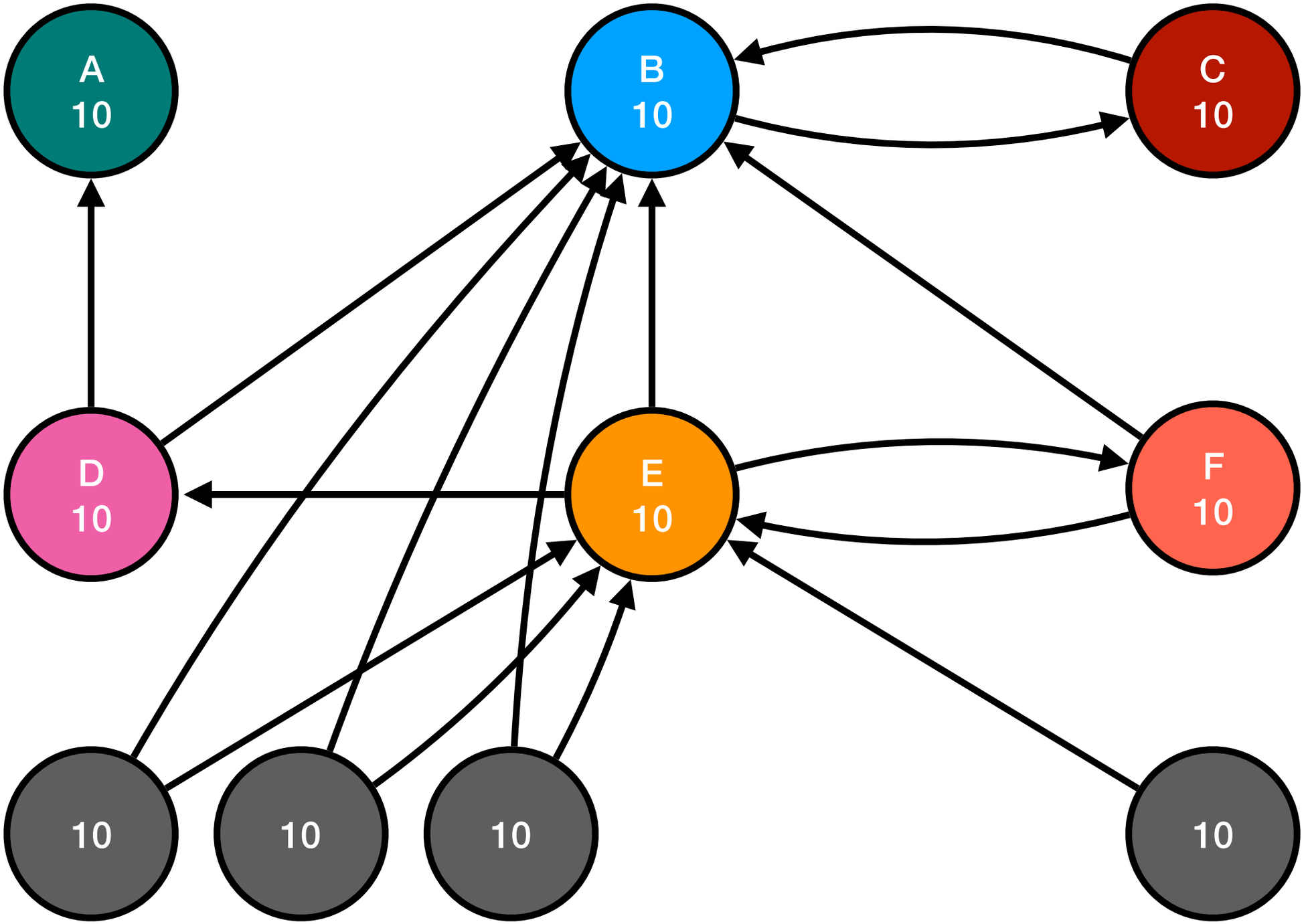
$$= \dots$$



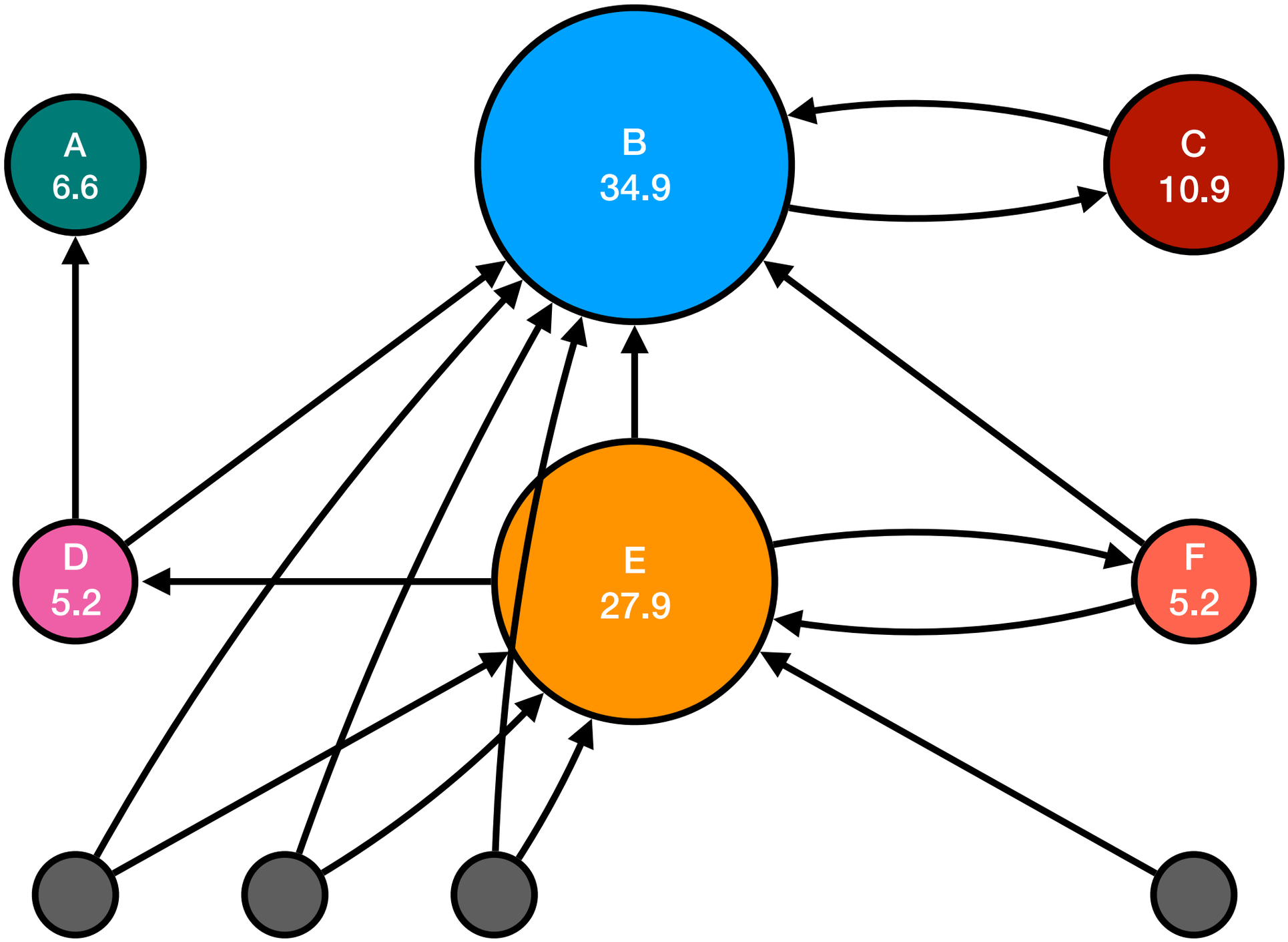
$$PR(x,0) = 1/3$$

$$PR(x,1) = \frac{1-0.85}{3} + 0.85 \cdot 1/3 = 1/3$$

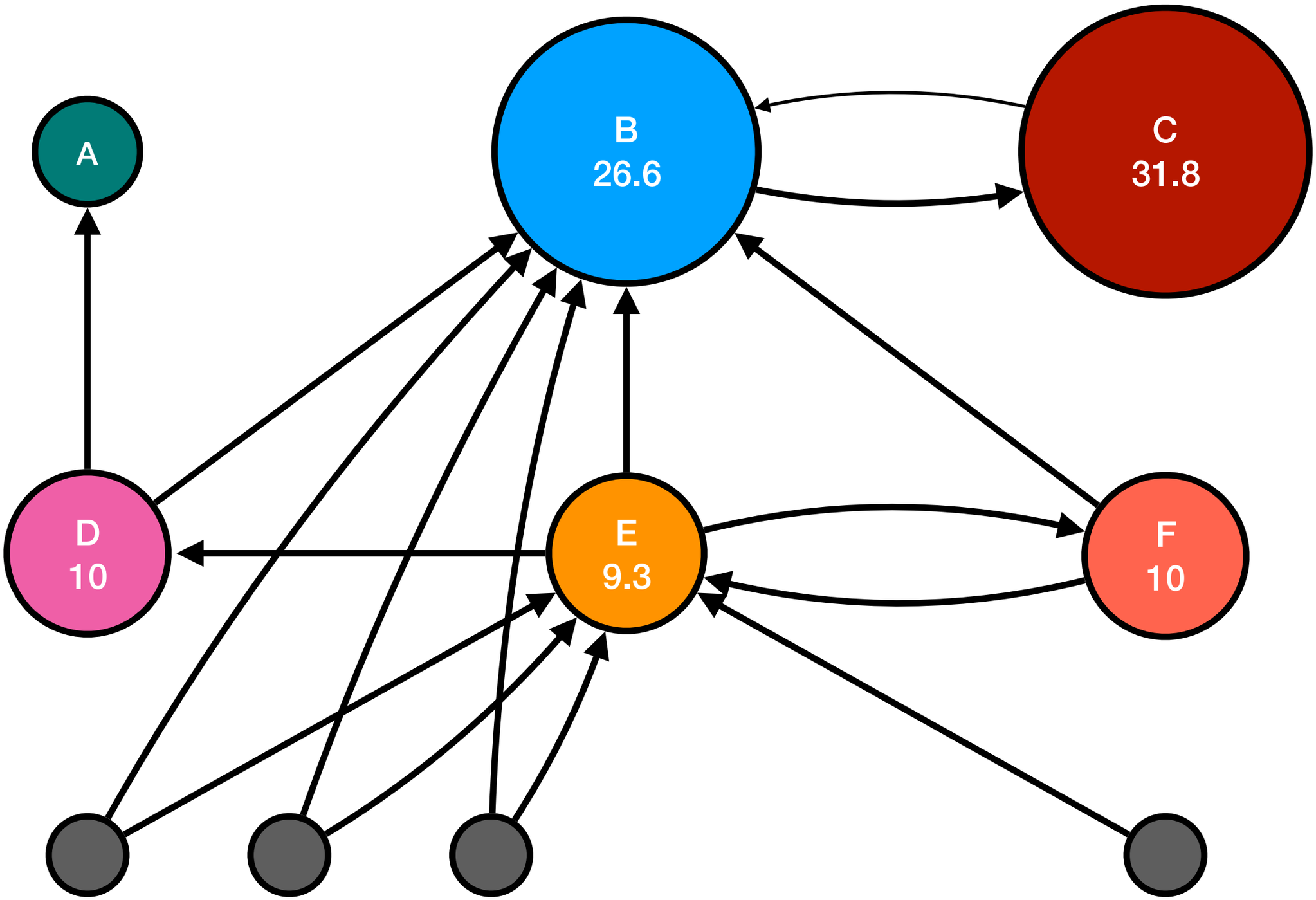
Pagerank (1/8)



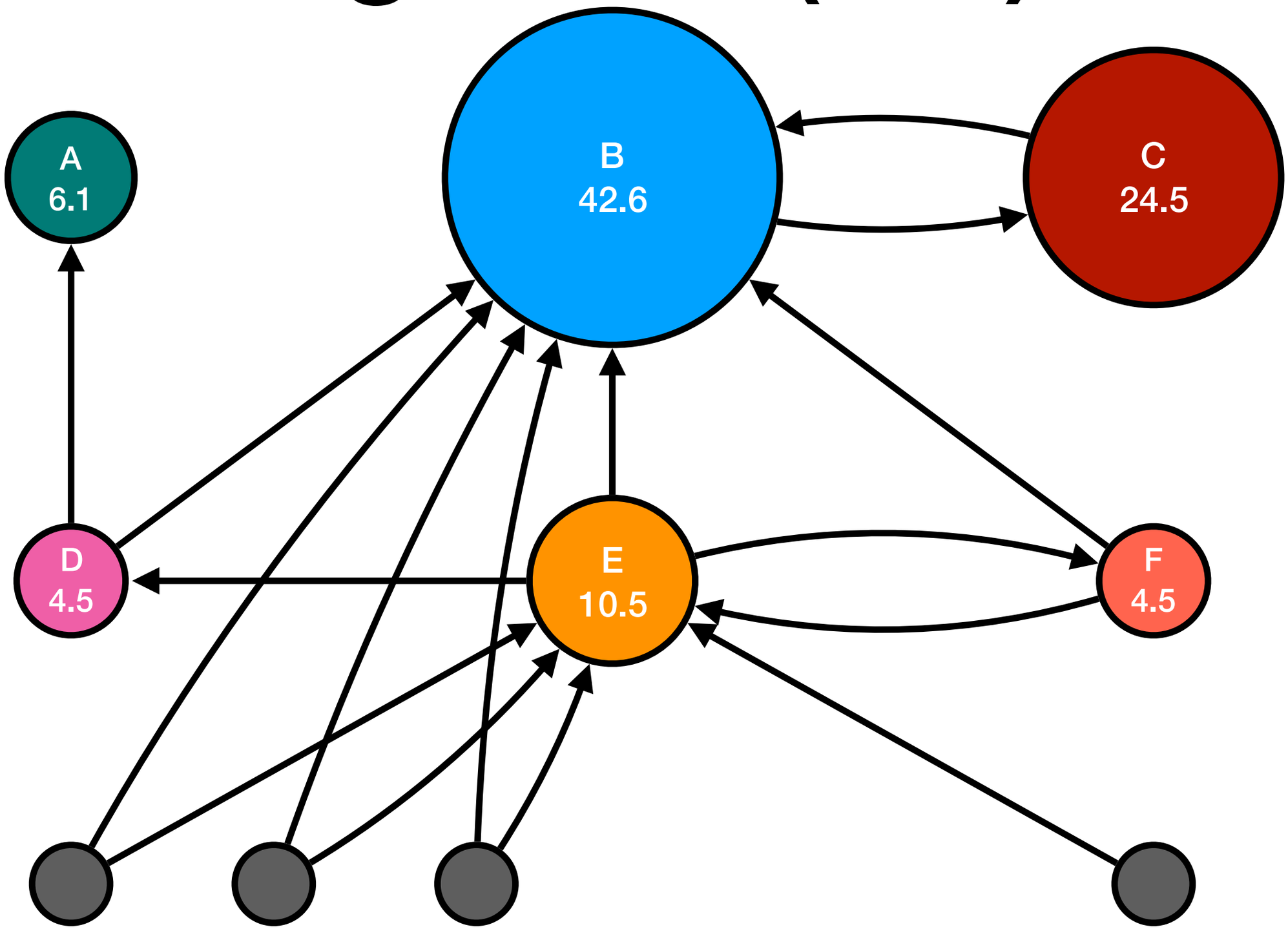
Pagerank (2/8)



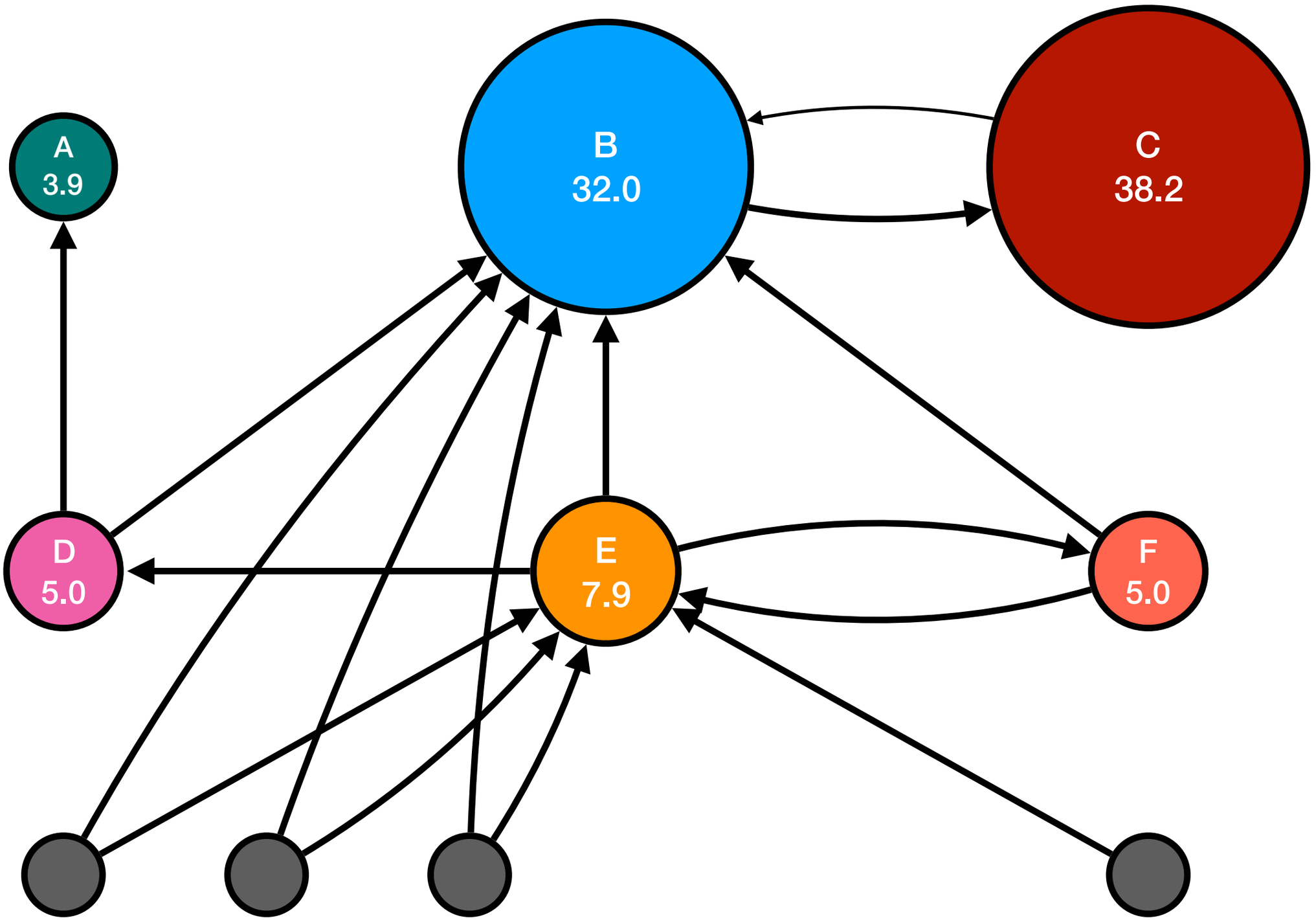
Pagerank (3/8)



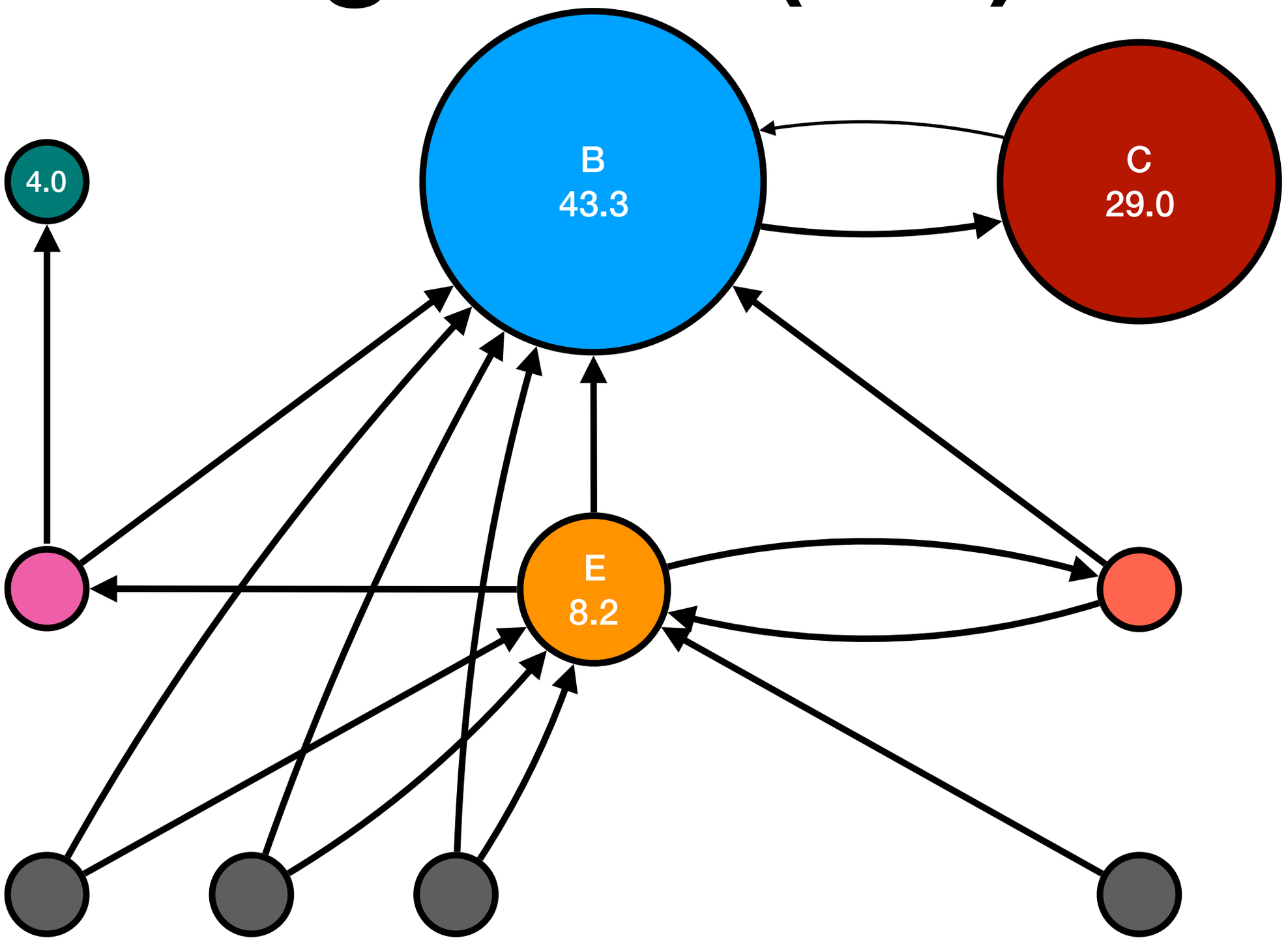
Pagerank (4/8)



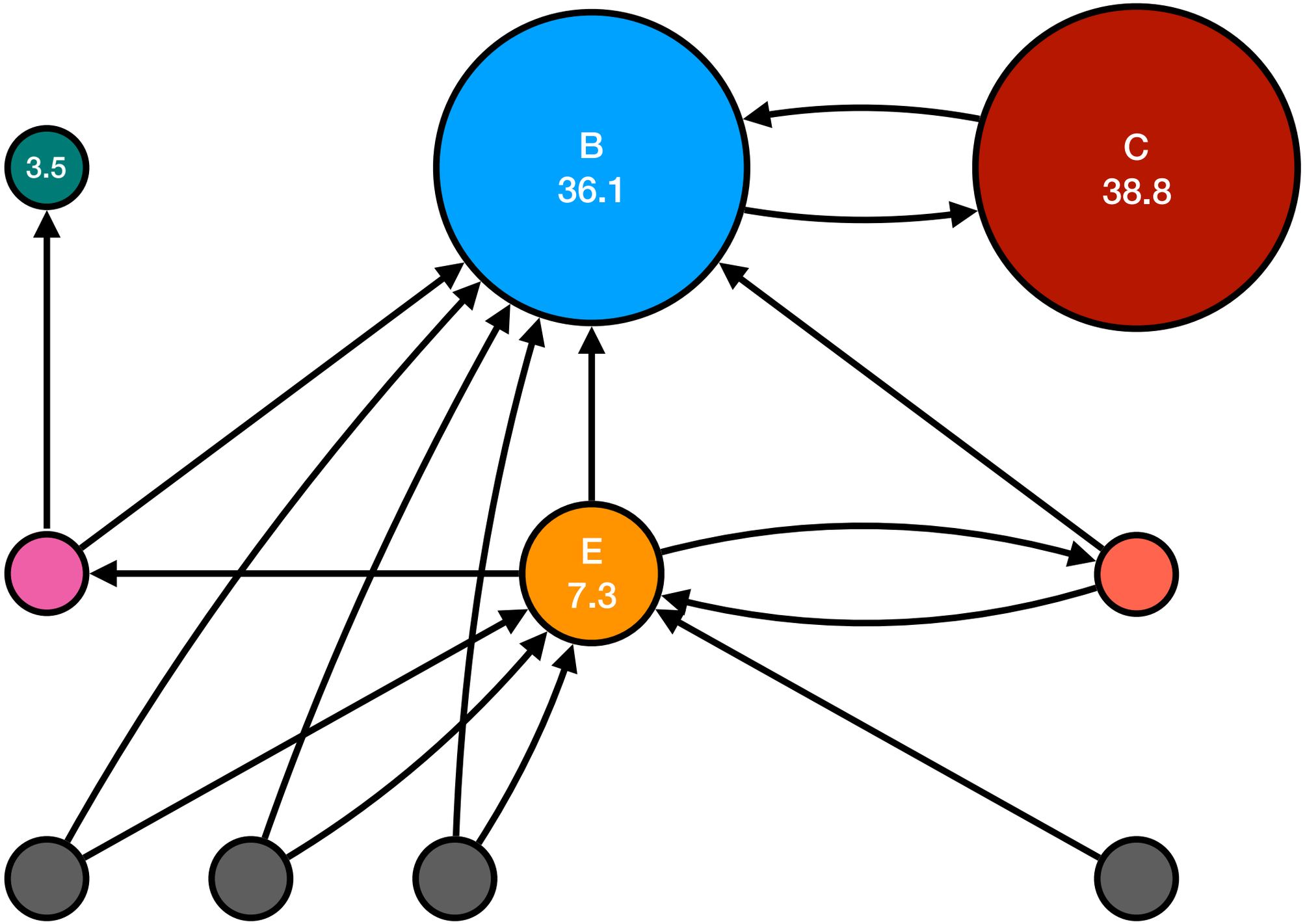
Pagerank (5/8)



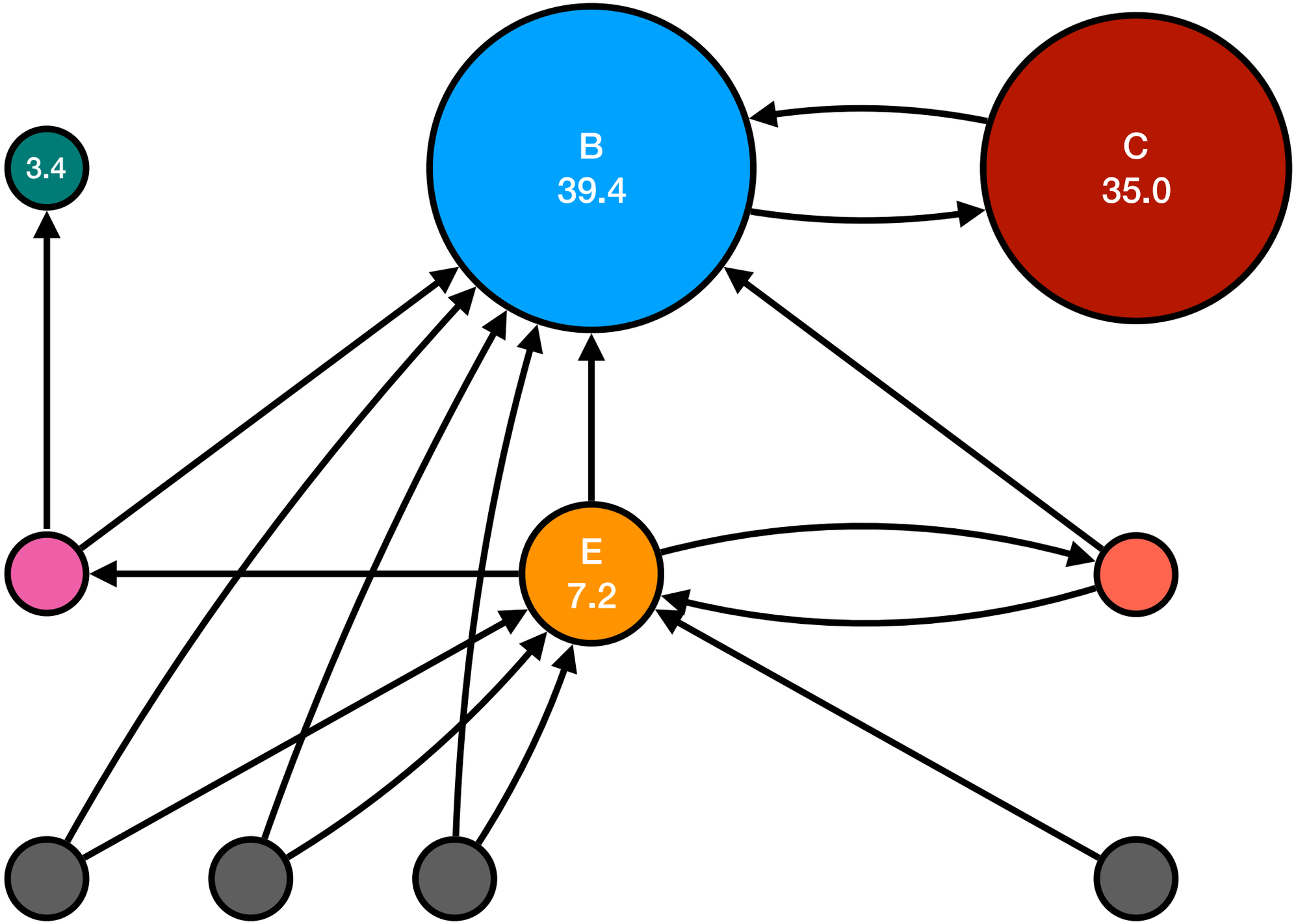
Pagerank (6/8)



Pagerank (7/8)



Pagerank (8/8)



Pagerank

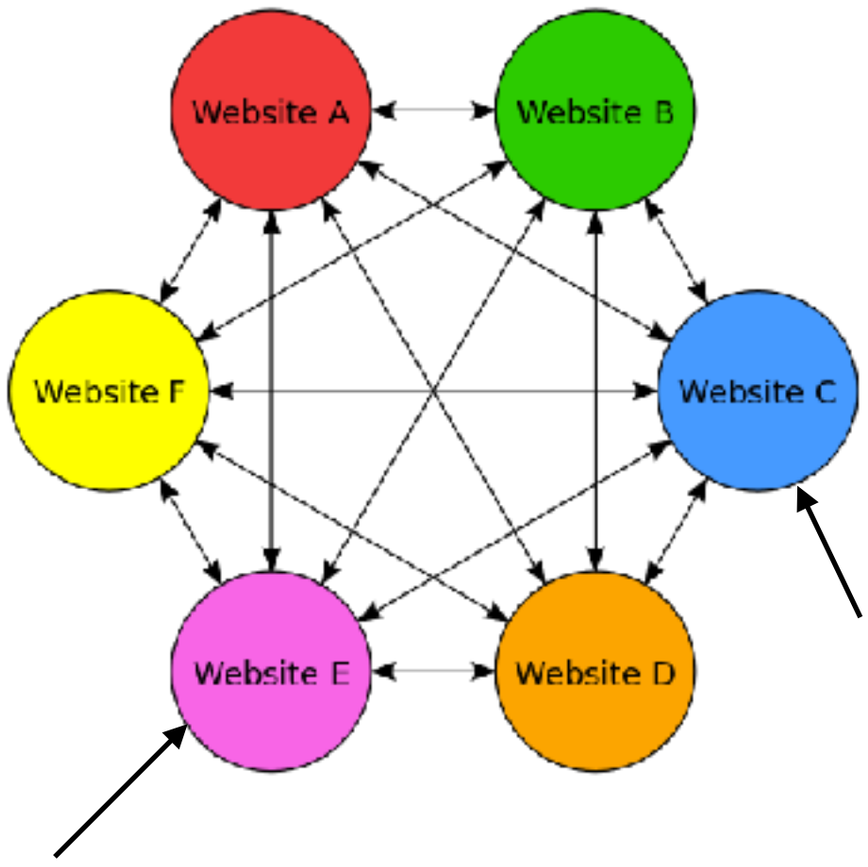
- 1998: 26 milioni di pagine, convergenza dopo ~40 iterazioni (27/09/1998, data “ufficiale” di lancio)
- 2000: 1 miliardo di pagine
- 2008: 1,000 miliardi di pagine
- 2014: 30,000 miliardi di pagine

Discussione

Cosa ne pensate di questo algoritmo?

Inizia una guerra

- Link farms
- Spam blogs
- Content farms
- Google bombing



Google

Web Images Groups News Froogle L

miserable failure

Web Results **1** - **10** of about 969,000

Biography of President George W. Bush
Biography of the president from the official White House web site.
www.whitehouse.gov/president/gwbbio.html - 29k - Cached - Similar pages
Past Presidents - Kids Only - Current News - President
More results from www.whitehouse.gov »

Welcome to MichaelMoore.com!
Official site of the gadfly of corporations, creator of the film Roger and Me and the television show The Awful Truth. Includes mailing list, message boar
www.michaelmoore.com/ - 35k - Sep 1, 2005 - Cached - Similar pages

BBC NEWS | Americas | '**Miserable failure**' links to Bush
Web users manipulate a popular search engine so an unflattering description to the president's page.
news.bbc.co.uk/2/hi/americas/3298443.stm - 31k - Cached - Similar pages

Google's (and Inktomi's) **Miserable Failure**
A search for **miserable failure** on Google brings up the official George W. Bush biography from the US White House web site. Dismissed by Google a
searchenginewatch.com/sereport/article.php/3296101 - 45k - Sep 1, 2005 -

La situazione oggi?

Google non pubblica più dettagli sul proprio algoritmo di ricerca

La *trasparenza* iniziale è andata perduta

- La rilevanza di una pagina viene calcolata a partire da 200 indicatori diversi
- L'algoritmo viene cambiato in continuazione
    - https://moz.com/google-algorithm-change

Ha dato origine ad un settore commerciale chiamato
SEO - Search engine optimization (72 G$ in 2017)

THE PERIODIC TABLE OF SEO SUCCESS FACTORS

Search engine optimization — SEO — seems like alchemy to the uninitiated. But there's a science to it. Below are some important "ranking factors" and best practices that can lead to success with both search engines and searchers.

ON-THE-PAGE FACTORS

These elements are in the direct control of the publisher

CONTENT

Cq	QUALITY	Are pages well written & have substantial quality content?
Cr	RESEARCH	Have you researched the keywords people may use to find your content?
Cw	WORDS	Do pages use words & phrases you hope they'll be found for?
Ce	ENGAGE	Do visitors spend time reading or "bounce" away quickly?
Cf	FRESH	Are pages fresh & about "hot" topics?
Vt	THIN	Is content "thin" or "shallow" & lacking substance?
Va	ADS	Is your content ad-heavy, especially "above the fold?"

HTML

Ht	TITLES	Do HTML title tags contain keywords relevant to page topics?
Hd	DESCRIPTION	Do meta description tags describe what pages are about?
Hh	HEADERS	Do headlines & subheads use header tags with relevant keywords?
Hs	STRUCTURE	Do pages use structured data to enhance listings?
Vs	STUFFING	Do you excessively use words you want pages to be found for?
Vh	HIDDEN	Do colors or design "hide" words you want pages to be found for?

ARCHITECTURE

Ac	CRAWL	Can search engines easily "crawl" pages on site?
Ad	DUPLICATE	Does site manage duplicate content issues well?
As	SPEED	Does site load quickly?
Au	URLS	Are URLs short & contain meaningful keywords to page topics?
Vc	CLOAKING	Do colors or design "hide" words you want pages to be found for?

ON-THE-PAGE SEO

CONTENT

- Cq +3 Quality
- Cr +3 Research
- Cw +2 Words
- Ce +2 Engage
- Cf +2 Fresh
- Vt -2 Thin
- Va -1 Ads

HTML

- Ht +3 Titles
- Hd +2 Description
- Hh +1 Headers
- Hs +1 Structure
- Vs -1 Stuffing
- Vh -1 Hidden

ARCHITECTURE

- Ac +3 Crawl
- Ad +2 Duplicate
- As +1 Speed
- Au +1 URLs
- Vc -3 Cloaking

OFF-THE-PAGE SEO

LINKS

- Lq +3 Quality
- Lt +2 Text
- Ln +1 Numbers
- Vp -3 Paid
- Vl -2 Spam

TRUST

- Ta +3 Authority
- Th +1 History
- Ti +1 Identity
- Vd -1 Piracy

SOCIAL

- Sr +2 Reputation
- Ss +1 Shares
- Ps +2 Social

PERSONAL

- Pc +3 Country
- Pl +3 Locality
- Ph +3 History

FACTORS WORK TOGETHER

All factors on the table are important, but those marked 3 carry more weight than 1 or 2. No single factor guarantees top rankings or success, but having several favorable ones increases the odds. Negative "violation" factors shown in red harm your chances.

OFF-THE-PAGE FACTORS

Elements influenced by readers, visitors & other publishers

LINKS

Lq	QUALITY	Are links from trusted, quality or respected web sites?
Lt	TEXT	Do links pointing at pages use words you hope they'll be found for?
Ln	NUMBER	Do many links point at your web pages?
Vp	PAID	Have you purchased links in hopes of better rankings?
Vl	SPAM	Have you created many links by spamming blogs, forums or other places?

TRUST

Ta	AUTHORITY	Do links, shares & other factors make site a trusted authority?
Th	HISTORY	Has site or its domain been around a long time, operating in same way?
Ti	IDENTITY	Does site use means to verify its identity & that of authors?
Vd	PIRACY	Has site been flagged for hosting pirated content?

SOCIAL

Sr	REPUTATION	Do those respected on social networks share your content?
Ss	SHARES	Do many share your content on social networks?

PERSONAL

Pc	COUNTRY	What country is someone located in?
Pl	LOCALITY	What city or local area is someone located in?
Ph	HISTORY	Has someone regularly visited your site or socially favored it?
Ps	SOCIAL	Have your friends socially favored the site?

Written By: Search Engine Land
Design By: COLUMN FIVE
Learn More: http://selnd.com/sestable

Incidenti - Google (2004)

Frank Weltner Presents...
This Scholarly Library of Facts about Domestic & Worldwide Zionist Criminality

Jew Watch

The Jew Watch Project Is The Internet's Largest Scholarly Collection of Articles on Zionist History
Free Educational Library for Private Study, Scholarship, Research & News About Zionism
We Reveal Zionist Banksters, News Falsifiers, PR Liars, Neocons, Subversives, Terrorists & Spies
The Jew Watch Project's 1.5 Billion Pages Served Demonstrate Our Focus on Professionalism
An Oasis of News for Americans Who Presently Endure the Hateful Censorship of Zionist Occupation

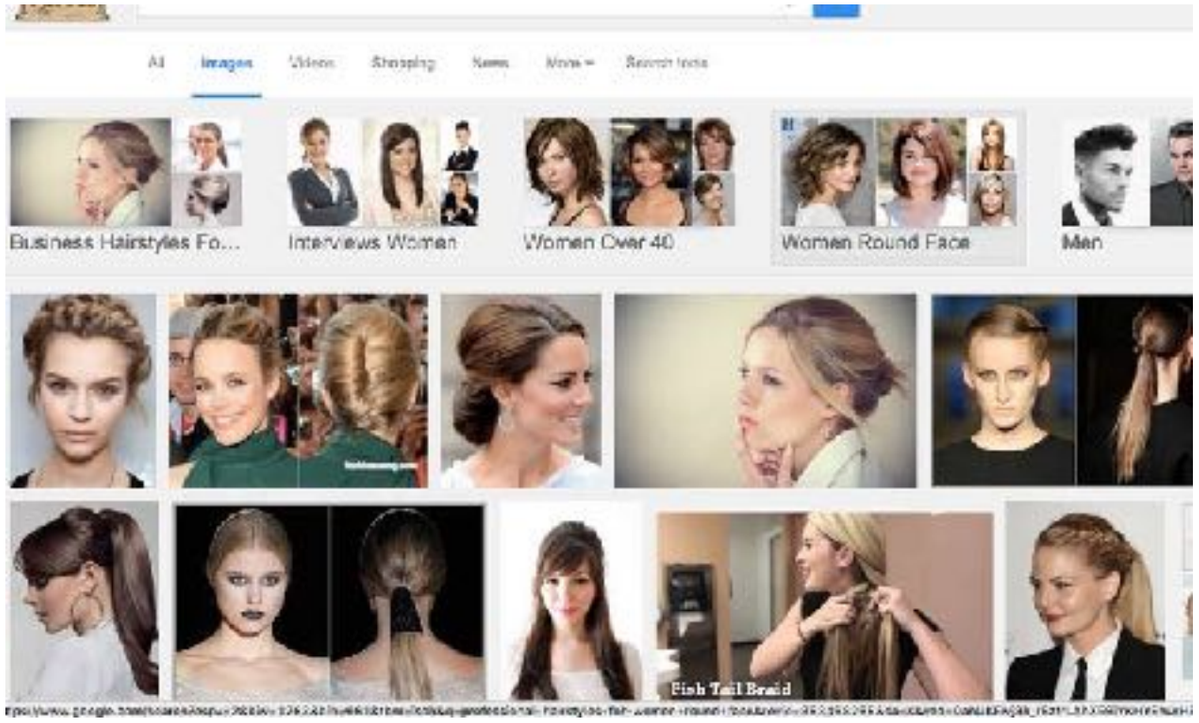
```
<a href="http://www.jewwatch.com/" rel="nofollow">
JewWatch sucks!
</a>
```

Incidenti - Google …

- 2006 - Google search & ads: “Foundem” (vertical price search)
- 2009 - Google images - Michelle Obama fake picture
- 2012 - Google autocomplete: “Bettina Wulff prostetuierte”

Incidenti - Google ...

- 2012 - Google ads: “Latanya Sweeney” vs “Tanya Smith”
- 2016 - Google images - “Unprofessional hairstyles for work” vs “Professional Hairstyles for work”



… ma non solo

2009 - Amazon top-sale list: libri gay-friendly rimossi

2011 - Twitter trending tags: #OccupyWallStreet

SIRI, where I can find an abortion clinic?

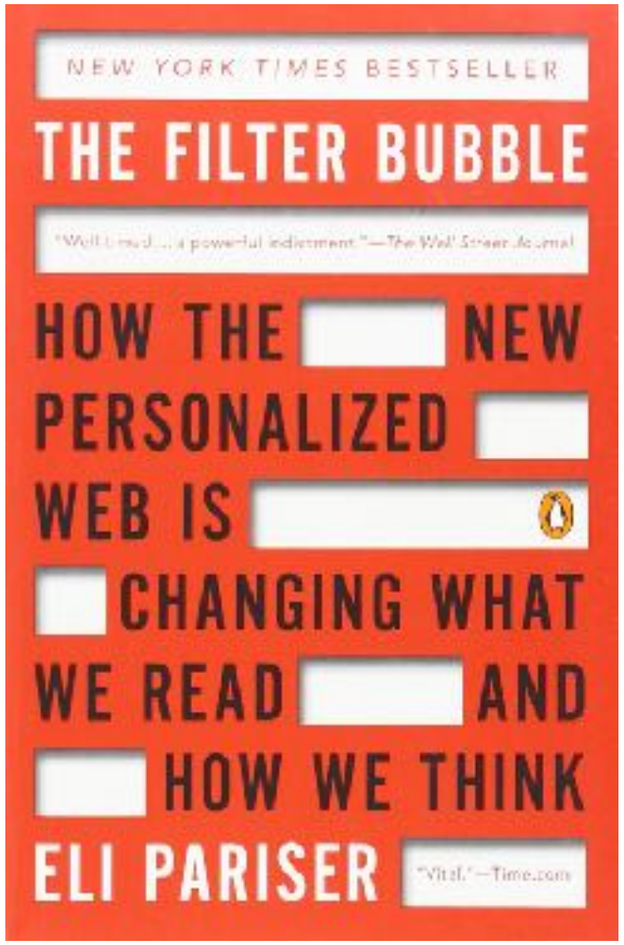
2012 - Target “You may also like…”: Prodotti per la gravidanza

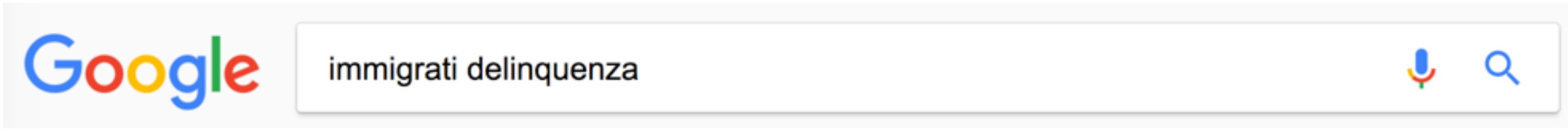
Filter bubble

I rischi della “personalizzazione”:

… *Google and Facebook are offering us too much candy, and not enough carrots* …

… *invisible algorithmic editing of the web may limit our exposure to new information and narrow our outlook*…





~~Facebook~~ Algorithm
Instagram

How Does Facebook Choose What To Show In News Feed?

$$\text{News Feed Visibility} =^{*} \underset{\text{Creator}}{C} \times \underset{\text{Post}}{P} \times \underset{\text{Type}}{T} \times \underset{\text{Recency}}{R}$$

Creator
Interest of the user in the creator

Post
This post's performance amongst other users

Type
Type of post (status, photo, link) user prefers

Recency
How new is the post

* This is a simplified equation. Facebook also looks at roughly 100,000 other high-personalized factors when determining what's shown.

Algoritmi ed Economia

TODAY IN PERSONAL JOURNAL

Goddesses in Bodices: The Latest in Taffeta

SPORTS Brooklyn Bridegrooms: Wedded to Victory in Baseball?

THE WALL STREET JOURNAL.

DOW JONES

* * * * *

MONDAY, JULY 8, 1889 - VOL. 1 NO. 1

★★ $0.02

Last week: DJ 12-STOCK AVG 87.71 ▼ 0.33 0.37% | 10-YR. TREASURY ▲ 1/32, yield 3.44% | OIL $0.91125 unch | POUND STERLING $4.87 unch | WHEAT $0.78625 ▲ $0.0375 | PORK $11.525 ▼ $0.075 | CORN $0.35625 unch

What's N

...volume ...at 96,520 shares.

C1

■ **As the U.S. Treasury** continues its massive program of bond buying, many investors fear the creation of a

Lunch, With a Side of Battle Plans

Thure Thulstrup

Traders gossip at a Wall Street eatery as the fight between the Consolidated and New York Stock Exchanges heats up

To read these stories in their entirety please go to WSJ.com/125.

Telegraph Steps Up Trading Speed

NEW YORK—The floor of the New York Stock Exchange is one of the most advanced technology centers on earth.

The perimeter of the cavernous "Main Board" room is lined with telephones, where clerks take orders cascading into lower Manhattan from around the nation. The clerks scrawl the orders onto trade slips, which they stuff into pneumatic tubes that blast the slips across the floor to the horseshoe-shaped trading stations. When the trade is executed, often in a minute or less, the broker's number is posted instantly overhead on the giant "annunciator boards" that line the walls of the exchange, signaling that he should pick up the telephone. Swarms of "telegraph

High Frequency Trading

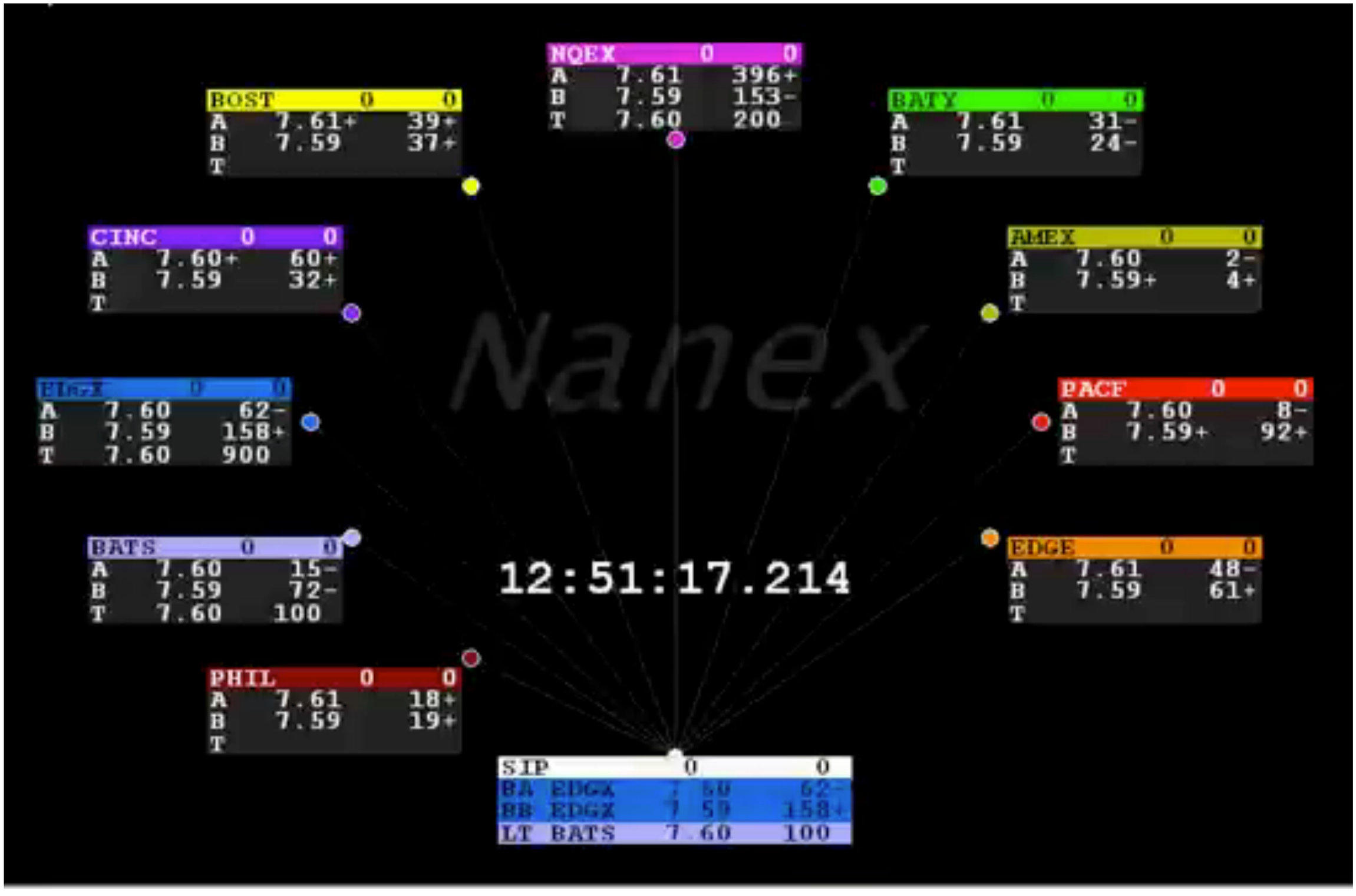
1983: NASDAQ introduce il primo mercato elettronico

2000: Tempo esecuzione scambi: < 10s

2010:

- Round-trip-delay: < 250µs (NASDAQ, colocation)
- Tempo di decisione: < 50µs

Slow Motion


NQEX 0 0
A 7.61 396+
B 7.59 153-
T 7.60 200
BOST 0 0
A 7.61+ 39+
B 7.59 37+
T
BATY 0 0
A 7.61 31-
B 7.59 24-
T
CINC 0 0
A 7.60+ 60+
B 7.59 32+
T
AMEX 0 0
A 7.60 2-
B 7.59+ 4+
T
EDGX 0 0
A 7.60 62-
B 7.59 158+
T 7.60 900
PACE 0 0
A 7.60 8-
B 7.59+ 92+
T
BATS 0 0
A 7.60 15-
B 7.59 72-
T 7.60 100
12:51:17.214
EDGE 0 0
A 7.61 48-
B 7.59 61+
T
PHIL 0 0
A 7.61 18+
B 7.59 19+
T
SIP 0 0
BA EDGX 7.60 62-
BB EDGX 7.59 158+
LT BATS 7.60 100

Incidenti

6 Maggio 2010:

- 09.00 Apre il Dow Jones
- 14.42 Il mercato perdeva 300 punti (su circa 10.000)
- 14.47 Il mercato perdeva 1000 punti
- 15.07 Il mercato tornava alla quotazione precedente, -300

1 Agosto 2012:

- Per un bug, Knight Capital ha perso $440 milioni in 45 minuti

Algorithmic ecology

- Predatori e prede
  - Algoritmi che "fanno liquidità”
  - Algoritmi che “prendono liquidità”
- Sistemi complessi
  - Comportamento emergente
  - Sistemi caotici

Algoritmi e Legge

AMMINISTRATIVO - NUOVE TECNOLOGIE

Diritto di accesso all'algoritmo, TAR Lazio apre nuovi scenari

TAR, Lazio-Roma, sez. III bis, sentenza 22/03/2017 n° 3769

di Michele Iaselli

Pubblicato il 17/05/2017

*Deve essere riconosciuto il diritto di accesso all'algoritmo del software di gestione di un procedimento amministrativo ogni qualvolta quest'ultimo possa configurarsi come atto amministrativo cd. informatico in senso stretto ai fini che interessano della sua piena ammissibilità in sede di accesso agli atti ai sensi dell'**art. 22 della legge n. 241 del 1990**.*

Algoritmi e Legge

Tribunale Ordinario di Milano

SI COMUNICA A:
Avv.PIANA CARLO
PIAZZA CASTELLO, 24
20121 MILANO MI
n. tel 0272023823 n. fax: 028690901

Sezione **01** -
Comunicazione di cancelleria
Tipo proced. Contenzioso
Numero di ruolo generale: **10847/2011**
Giudice: PADOVA MARIA LUISA
Data prossima udienza: Ore:

Parti nel procedimento

Attore principale GOOGLE INC [redacted]

Convenuto principale [redacted]

Avv. MONTANARO ROMOLO

Avv. PIANA CARLO

Algoritmi e Sociologia

Memphis Flyer

Blue Crush Continues To Help MPD Combat Crime

To celebrate the Flyer's 25th year, we'll be using this space each week to look back on stories from past issues.

In 2005, former Memphis Police Director Larry Godwin introduced a plan to crush crime one statistic at a time.

Called Blue CRUSH, the data-driven initiative uses information collected from Memphis Police Department (MPD) reports to determine local crime hotspots. Aggravated assaults, nondomestic violence, robbery, burglary, and vehicle theft are among the crimes targeted.

Discussione

Ma allora gli algoritmi sono brutti e cattivi?

Computational Science

Related fields [edit]

- Bioinformatics
- Cheminformatics
- Chemometrics
- Computational archaeology
- Computational biology
- Computational chemistry
- Computational materials science
- Computational economics
- Computational electromagnetics
- Computational engineering
- Computational finance
- Computational fluid dynamics
- Computational forensics
- Computational geophysics
- Computational history
- Computational informatics
- Computational intelligence
- Computational law
- Computational linguistics
- Computational mathematics
- Computational mechanics
- Computational neuroscience
- Computational particle physics
- Computational physics
- Computational sociology
- Computational statistics
- Computational sustainability
- Computer algebra
- Computer simulation
- Financial modeling
- Geographic information system (GIS)
- High-performance computing
- Machine learning
- Network analysis
- Neuroinformatics
- Numerical linear algebra
- Numerical weather prediction
- Pattern recognition
- Scientific visualization
- Simulation

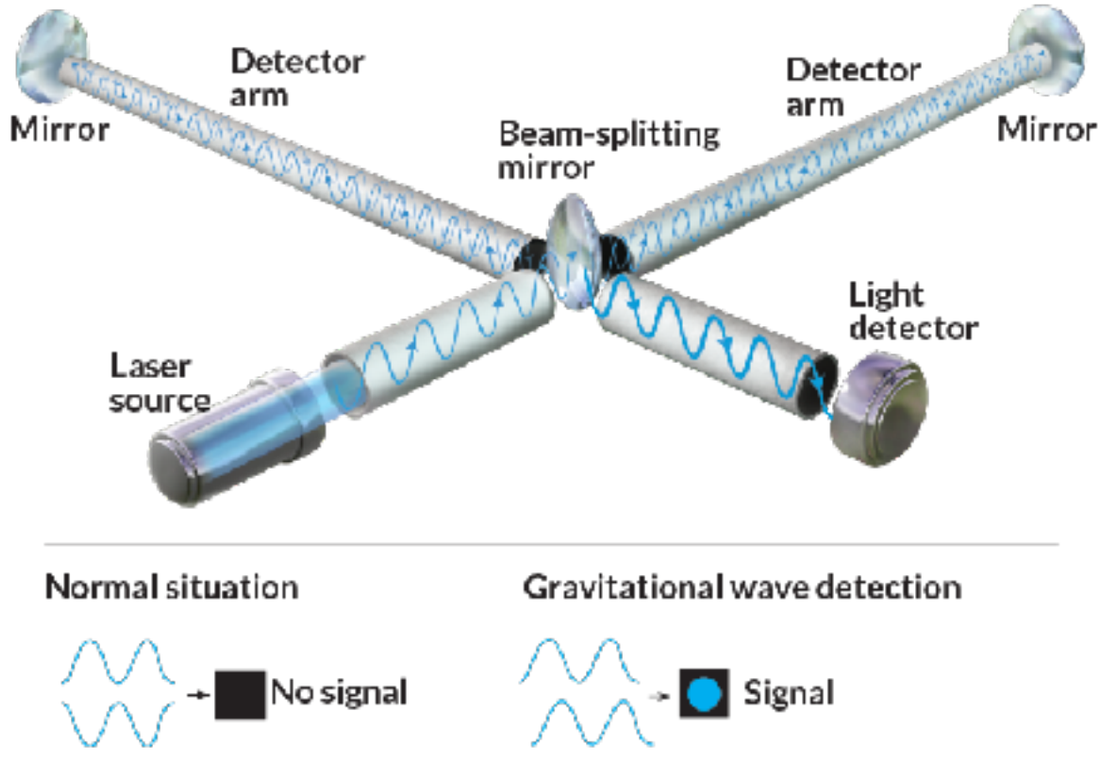
Esempi - Fisica

PHYSICAL REVIEW D
covering particles, fields, gravitation, and cosmology

Highlights Recent Accepted Authors Referees Search Press About

Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement

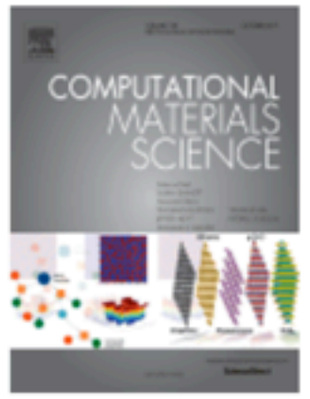
Benjamin J. Owen and B. S. Sathyaprakash
Phys. Rev. D **60**, 022002 – Published 22 June 1999



Esempi - Ingegneria

Computational Materials Science

Volume 25, Issue 3, November 2002, Pages 478-492



First-principles computation of material properties: the ABINIT software project

X. Gonze [a], J.-M. Beuken [a], R. Caracas [a], F. Detraux [a], M. Fuchs [a], G.-M. Rignanese [a], L. Sindic [a], M. Verstraete [a], G. Zerah [b], F. Jollet [b], M. Torrent [b], A. Roy [b], M. Mikami [c], Ph. Ghosez [d], J.-Y. Raty [d], D.C. Allan [e]

Computational Biology

DNA: Una stringa di molecole chiamate basi
(Adenina, Citosina, Guanina, Timina)
Esempio: AGACCCTTAA

Problema: date due sequenze di DNA, trovare quanto siano "simili"
Esempio: AAAATTGA, TAACGATA

Sottosequenza comune massimale: trovare la più lunga sottosequenza (non necessariamente contigua) presente in entrambe le sequenze
Esempio: AAAATTGA, TAACGATA

Una possibile soluzione

- Date due sequenze S_1, S_2:
- *max* = sottosequenza vuota
- Per ogni sottosequenza *x* in S_1:
  - Se *x* è sottosequenza di S_2
    - Se *x* è più lunga di max:
      - *max* = *x*

- Quante sottosequenze esistono?

  2^n

```
AGTC

AGTC 1111
AGT  1110
AG C 1101
AG   1100
A TC 1011
A T  1010
A  C 1001
A    1000
 GTC 0111
 GT  0110
 G C 0101
 G   0100
  TC 0011
  T  0010
   C 0001
     0000
```

Descrizione matematica della soluzione ottima

Prefisso (Prefix)

Data una sequenza P composta dai caratteri $p_1 p_2 \ldots p_n$, $P(i)$ denota il prefisso di P dato dai primi i caratteri, i.e.:

$$P(i) = p_1 p_2 \ldots p_i$$

Esempi

- $P =$ "`ABDCCAABD`"
- $P(0) = \emptyset$ (sequenza vuota)
- $P(3) =$ "`ABD`"
- $P(6) =$ "`ABDCCA`"

Descrizione matematica della soluzione ottima

Goal

Date due sequenze P e T di lunghezza n e m, scriviamo una formula ricorsiva $LCS(P(i), T(j))$ che restituisca la LCS dei prefissi $P(i)$ e $T(j)$.

$$LCS(P(i), T(j)) = \begin{cases} ? & \text{Caso base} \\ ? & \text{Casi ricorsivi} \end{cases}$$

Casi ricorsivi

Caso 1

Considerate due prefissi $P(i)$ e $T(j)$ tali per cui l'ultimo loro carattere coincide: $p_i = t_j$. Come calcolereste la LCS di $P(i)$ e $T(j)$?

- Esempio: $P(i) =$ "ALBERTO", $T(j) =$ "PIERO"

Casi ricorsivi

Caso 1

Considerate due prefissi $P(i)$ e $T(j)$ tali per cui l'ultimo loro carattere coincide: $p_i = t_j$. Come calcolereste la LCS di $P(i)$ e $T(j)$?

- Esempio: $P(i) =$ "ALBERTO", $T(j) =$ "PIERO"

Soluzione

$$LCS(P(i), T(j)) = LCS(P(i-1), T(j-1)) \oplus p_i$$

dove $\oplus$ è l'operatore di concatenazione.

- $LCS($"ALBERTO","PIERO"$) = LCS($"ALBERT","PIER"$) \oplus$ "O"

Casi ricorsivi

Caso 2

Considerate due prefissi $P(i)$ e $T(j)$ tali per cui l'ultimo loro carattere è differente: $p_i \neq t_j$. Come calcolereste la LCS di i e j?

- Esempio:$P(i) =$ "ALBERT", $T(j) =$ "PIER"

Soluzione

Casi ricorsivi

Caso 2

Considerate due prefissi $P(i)$ e $T(j)$ tali per cui l'ultimo loro carattere è differente: $p_i \neq t_j$. Come calcolereste la LCS di i e j?

- Esempio:$P(i) =$ "ALBERT", $T(j) =$ "PIER"

Soluzione

$$LCS(P(i), T(j)) = longest(LCS(P(i-1), T(j)), LCS(P(i), T(j-1))$$

- $LCS($"ALBERT","PIER"$) =$
  $longest(LCS($"ALBER","PIER"$), LCS($"ALBERT","PIE"$))$

Casi base

Casi base

Qual è la più lunga sottosequenza di $P(i)$ e $T(j)$, quando uno dei prefissi è vuoto, i.e. se $i = 0$ **or** $j = 0$?

- Esempio: $P(i) =$ "ALBERTO", $T(0) = \emptyset$

Soluzione

$$LCS(P(i), T(0)) = \emptyset$$

- $LCS(\text{"ALBERTO"}, \emptyset) = \emptyset$

La formula completa

$$LCS(P(i),T(j)) = \begin{cases} \emptyset & i = 0 \textbf{ or } j = 0 \\ LCS(P(i-1),T(j-1)) \oplus p_i & i > 0 \textbf{ and } j > 0 \textbf{ and } p_i = t_j \\ longest(LCS(P(i-1),T(j)), & \\ \quad LCS(P(i),T(j-1)) & i > 0 \textbf{ and } j > 0 \textbf{ and } p_i \neq t_j \end{cases}$$

Dimostrazione

Il fatto che la formula sia corretta dovrebbe essere provato. La dimostrazione è per assurdo.

Mi devo fermare qui…

Opportunamente progettato (programmazione dinamica), l'algoritmo può operare in un tempo che è quadratico nella dimensione delle sequenze

- Con un input di dimensione n=100
  - Algoritmo 2^n: 40 milioni di miliardi di anni
  - Algoritmo n^2: 0.2 ms

Bio-inspired Computation

- Epidemic protocols - virus
- Deep learning – neuroni e cervello
- Evolutionary Computation – evoluzione
- Swarm Intelligence – comportamento di gruppo
- Immunocomputing – sistema immunitario
- Artificial Life – vita artificiale
- Membrane Computing – celle e membrane
- Amorphous Computing - morfogenesi

Algoritmi genetici

Algoritmi che sono ispirati dall'evoluzione naturale, che sono basati su quattro elementi principali:

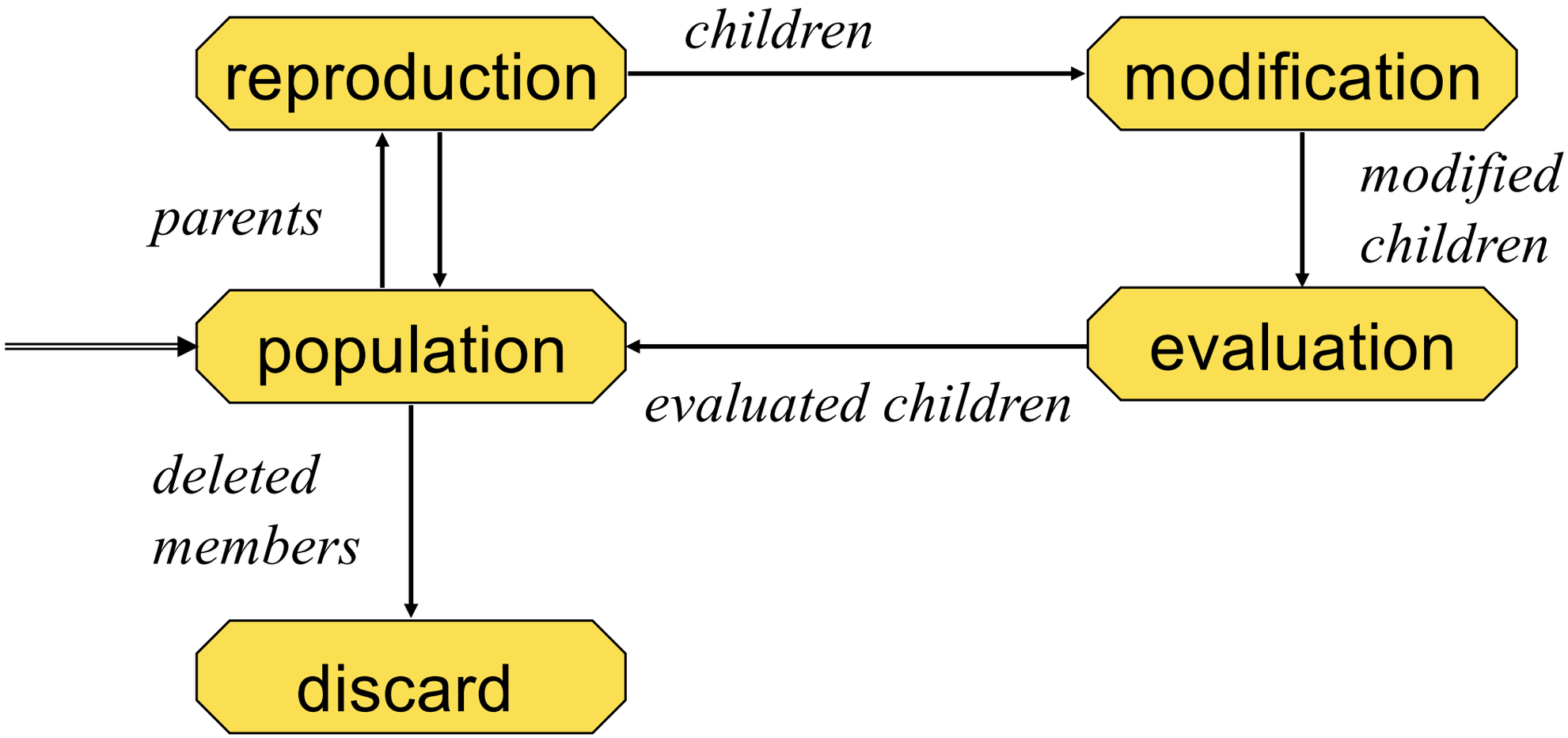
- Gruppi di individui - Popolazione
- Sorgente di variazione - Operatori genetici
- Fitness riproduttiva - Fitness
- Survival of the Fittest - Selezione

Come funziona?

- Per tentativi e errori
- "Ricette" per scegliere il prossimo tentativo

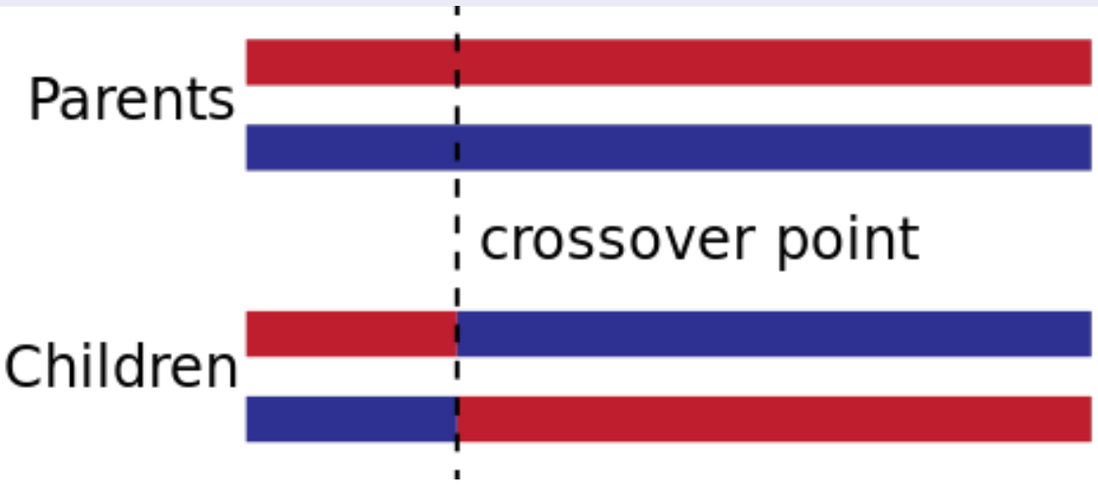
Il ciclo riproduttivo degli algoritmi genetici

The GA Cycle of Reproduction

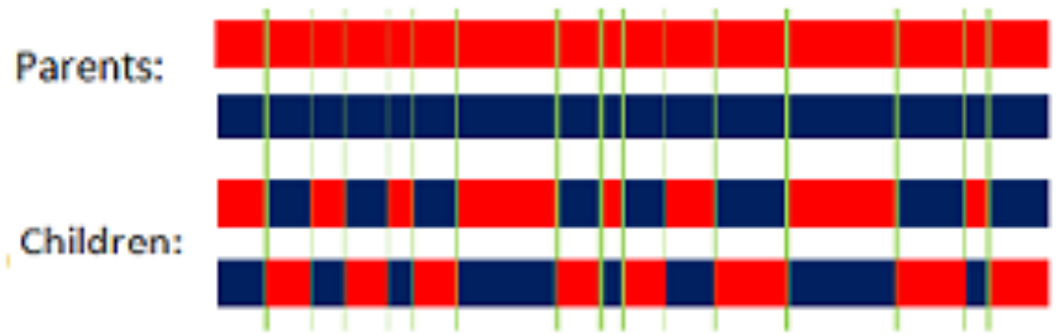


Operatori genetici

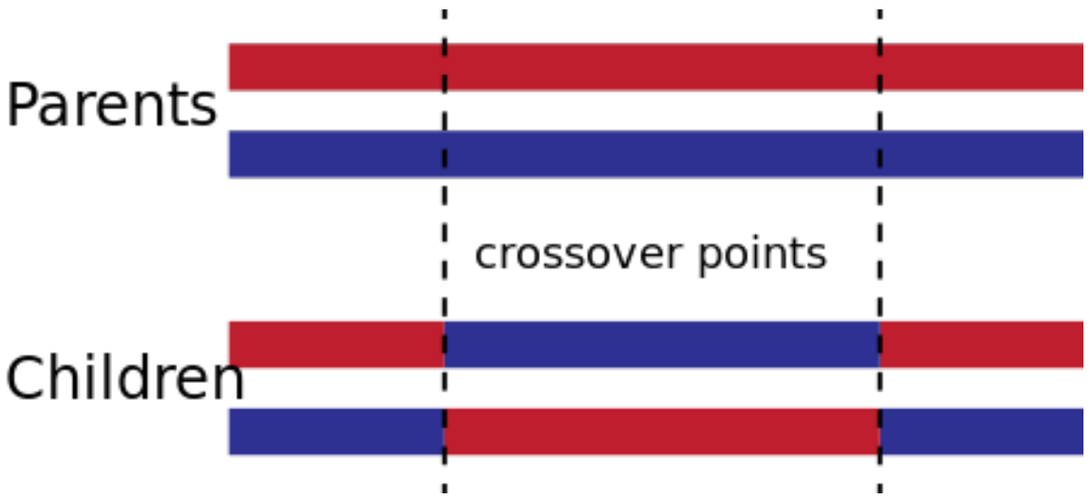
Single-point crossover



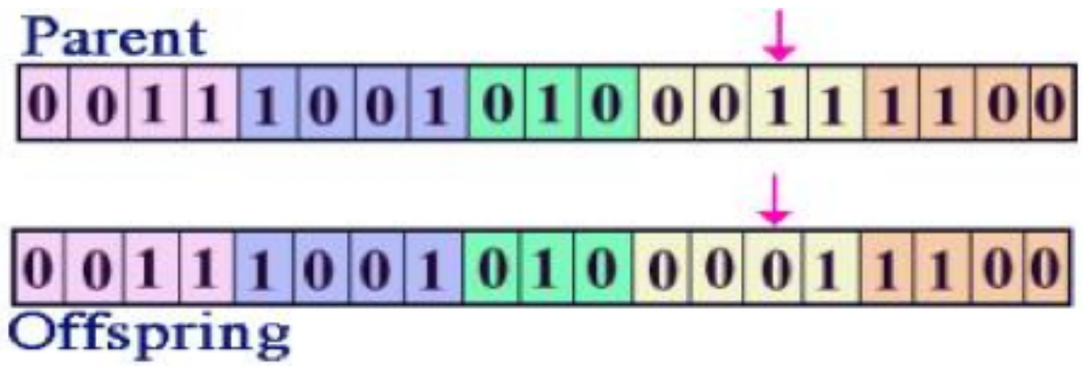
Uniform Crossover



Two-point crossover



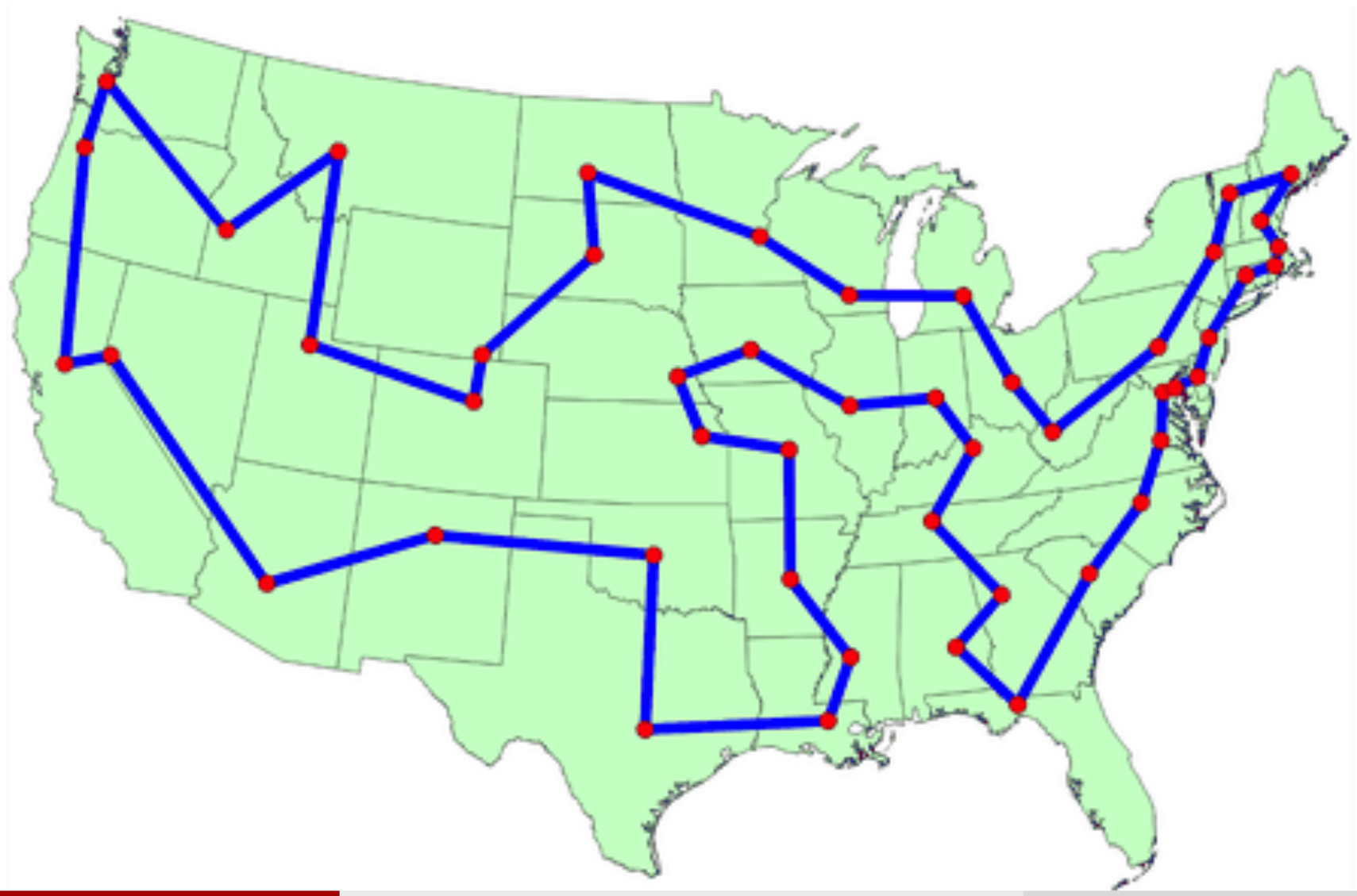
Mutatation



Traveling Salesman Problem

Problema

Trovare un tour di un insieme di città tali per cui ogni città viene visitata una volta sola, e la distanza totale è minimizzata.



Rappresentazione

In questo esempio, si utilizza una rappresentazione basata su una permutazione di città

1) Berlin
2) Munich
3) Stuttgart
4) Wesbaden
5) Cologne
6) Hanover
7) Dusseldorf
8) Breme

Esempio

Lista 1: `(3 5 7 2 1 6 4 8)`
Lista 2: `(2 5 7 6 8 1 3 4)`

Mutazioni

Problema

Cambiare un solo elemento nella permutazione darebbe una soluzione non ammissibile (città replicate)

Esempio

(3 5 7 2 1 6 4 8) (Prima)
(3 5 5 2 1 6 4 8) (Dopo)

Soluzione

- Si scelgono due elementi casualmente
- Si muove il secondo subito dopo il primo, spostando il resto
- Questo preserva l'ordine e le informazioni di adiacenza

Esempio

(3 5 7 2 1 6 4 8) (Prima)
(3 5 7 6 2 1 4 8) (Dopo)

Crossover

Problema

Operazioni di crossover "standard" potrebbero generare città replicate

Esempio

(3 5 7 2 1 6 4 8)
(2 5 7 6 8 1 3 4)
può diventare:
(3 5 7 2 8 1 3 4
(2 5 7 6 1 6 4 8)

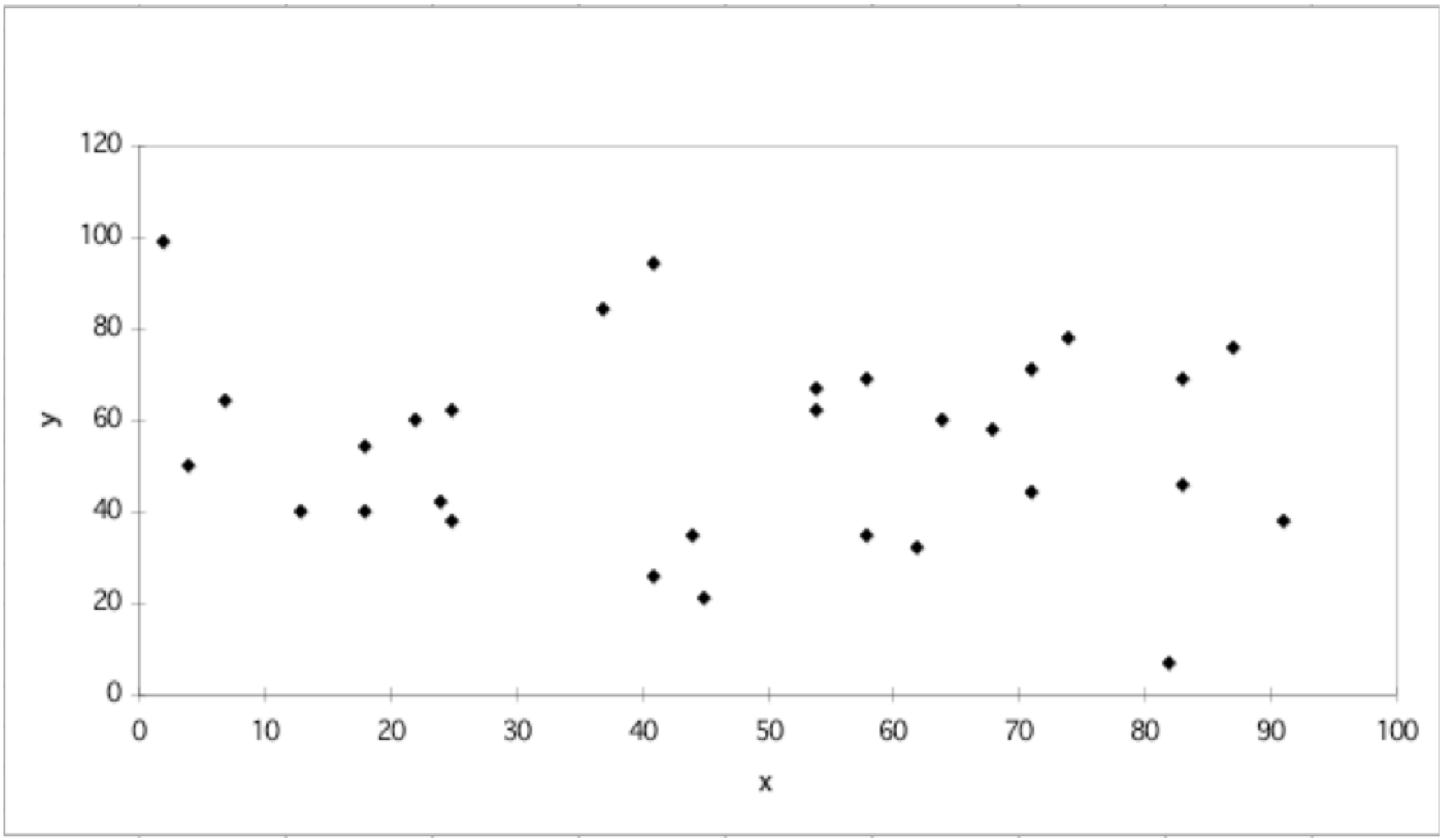
Soluzione

- Seleziona una parte arbitraria del primo genitore
- Copia questa parte nel primo figlio
- Aggiungi i numeri mancanti, nell'ordine in cui compaiono nel secondo genitore

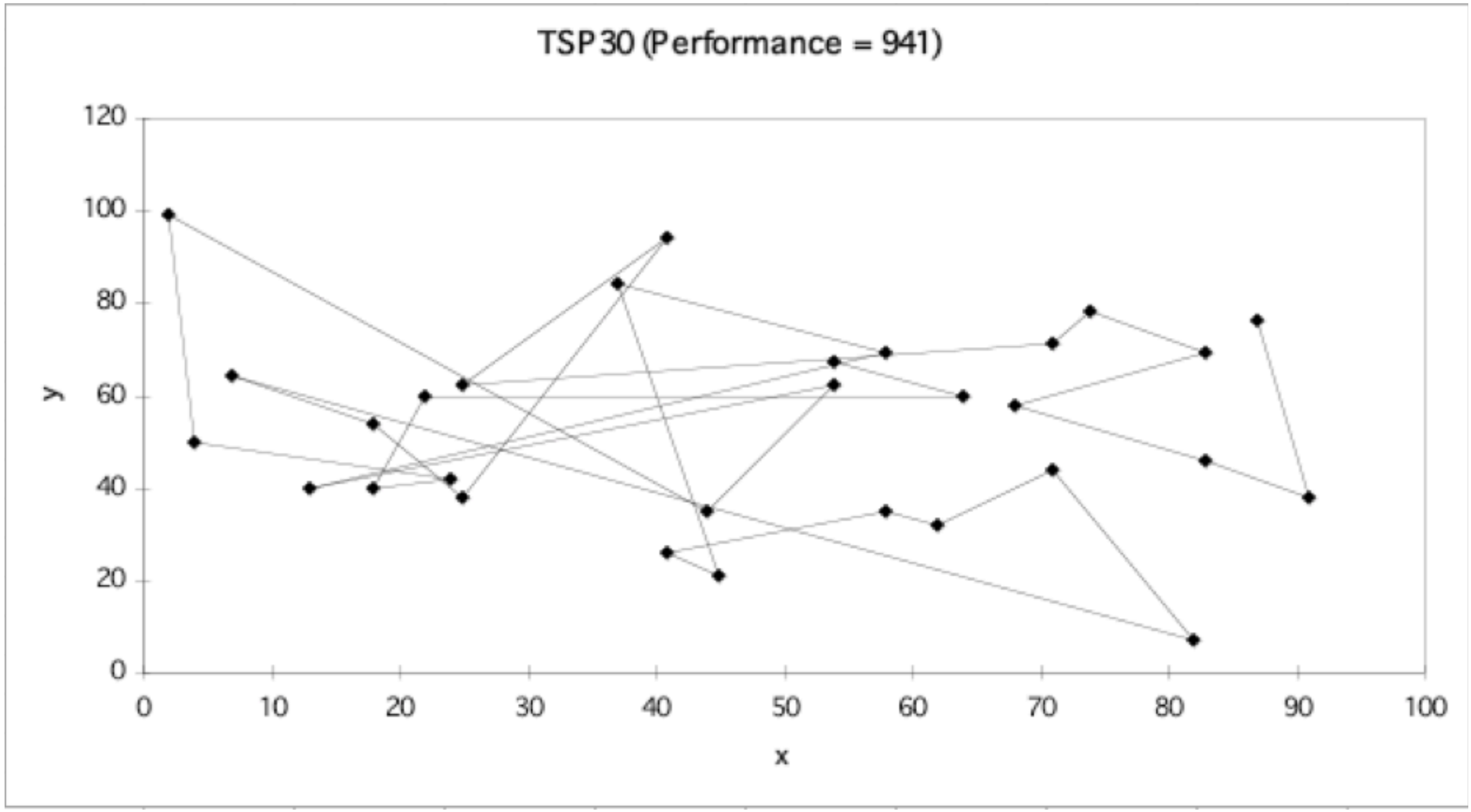
Esempio

(1 2 3 4 5 6 7 8)
(8 5 7 2 1 6 4 3)
può generare:
(8 7 3 4 5 6 2 1)

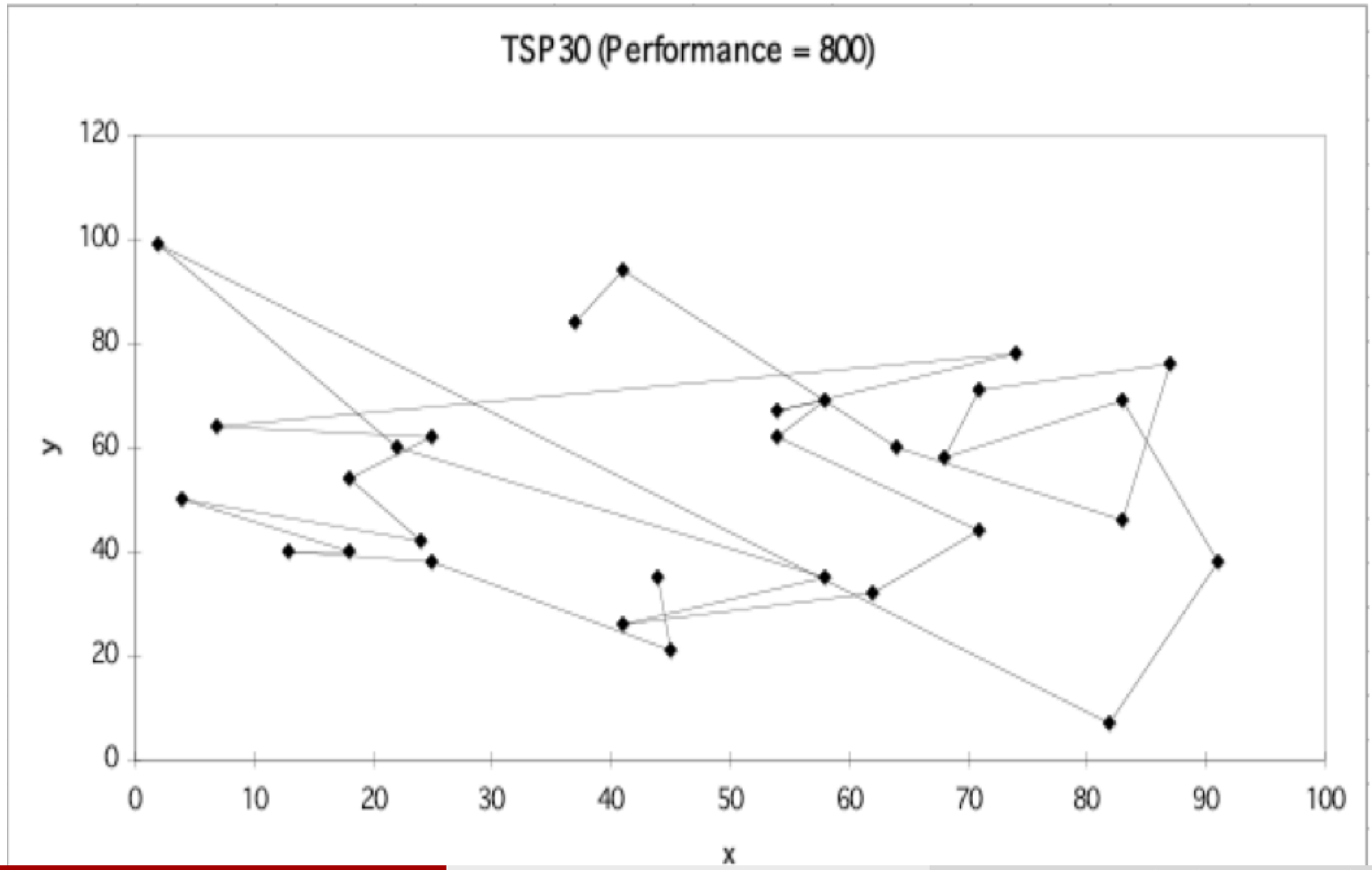
Results



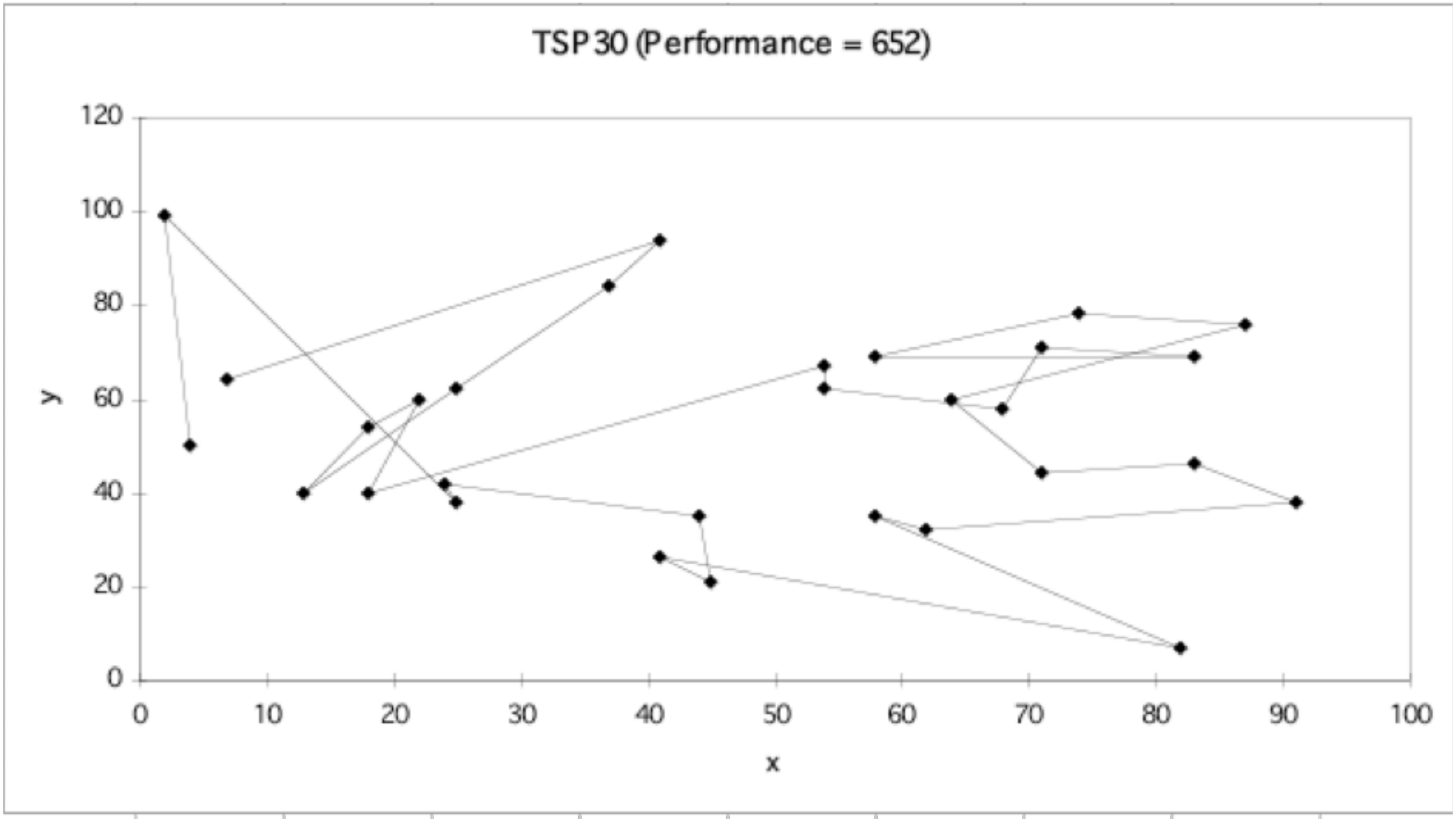
Results


TSP 30 (Performance = 941)
120
100
80
60
40
20
0
y
0
10
20
30
40
50
60
70
80
90
100
x

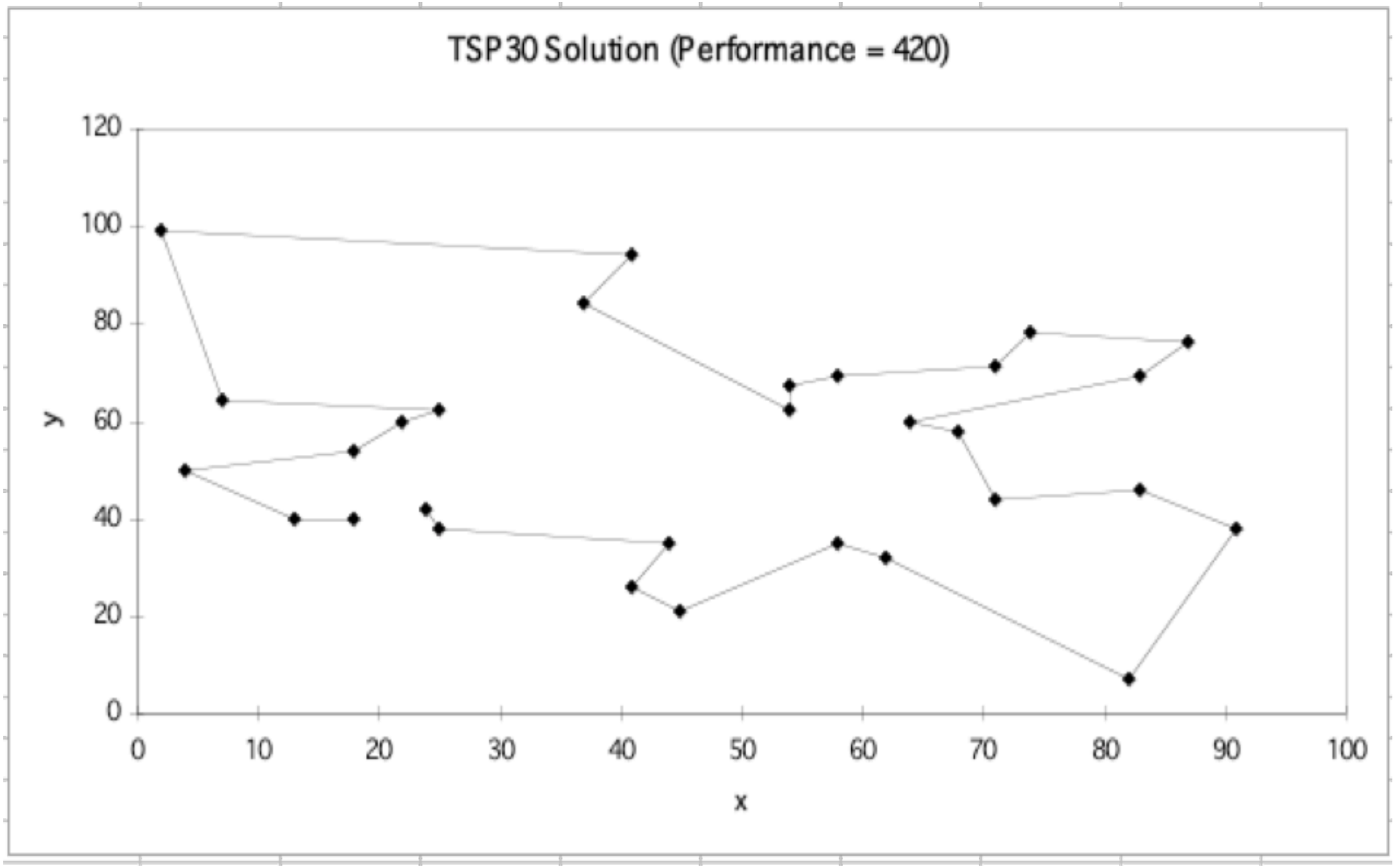
Results


TSP30 (Performance = 800)
120
100
80
60
40
20
0
y
0
10
20
30
40
50
60
70
80
90
100
x

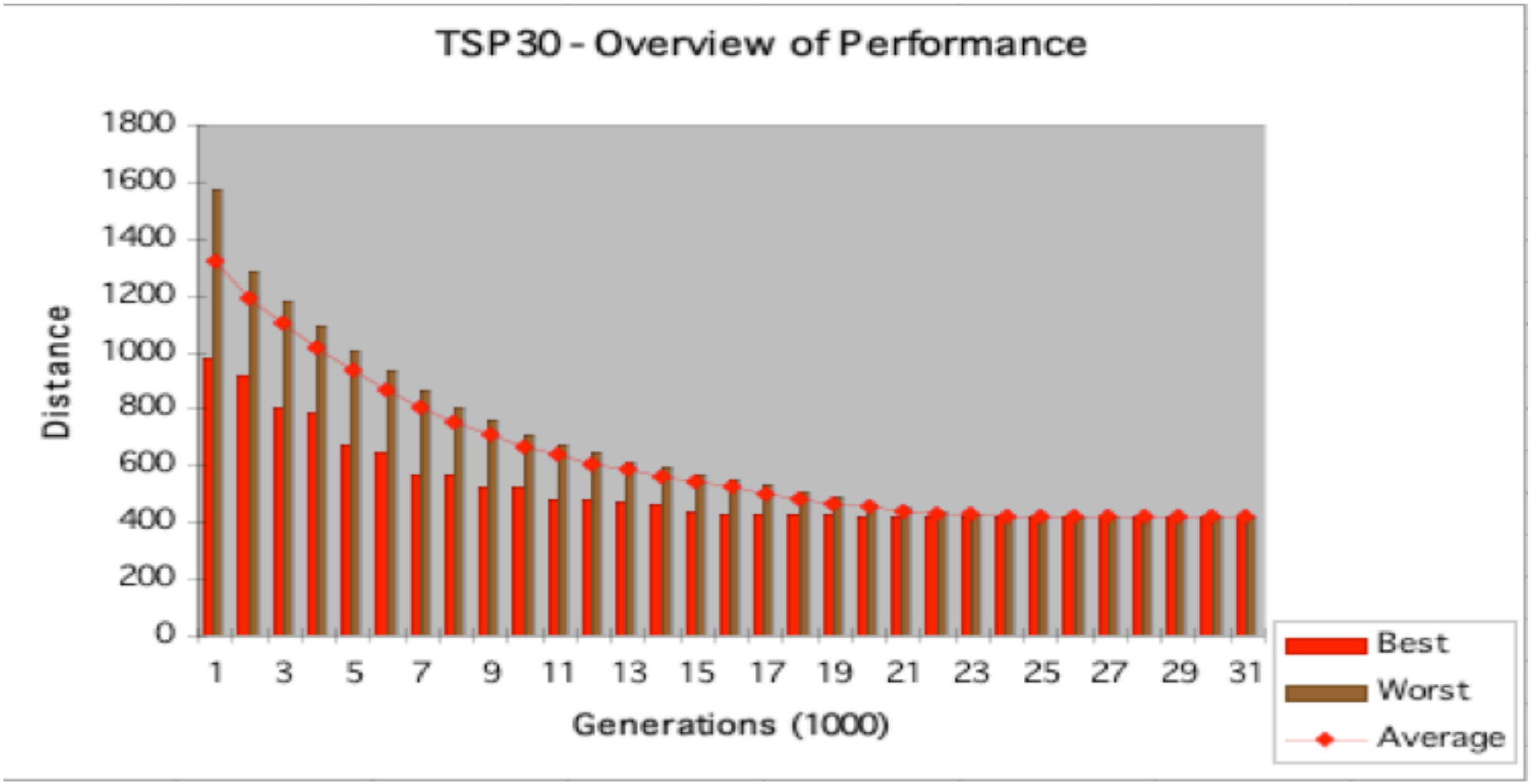
Results


TSP 30 (Performance = 652)
120
100
80
60
40
20
0
y
0
10
20
30
40
50
60
70
80
90
100
x

Results


TSP 30 Solution (Performance = 420)
120
100
80
60
40
20
0
y
0
10
20
30
40
50
60
70
80
90
100
x

Performance



Come cittadini, comprendere il ruolo degli algoritmi è fondamentale per affrontare il mondo moderno

Come futuri lavoratori, le posizioni più interessanti richiederanno competenze informatiche molto maggiori di ECDL

Da un grande potere (algoritmico) derivano grandi responsabilità!