ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΑΝΟΙΚΤΑ ακαδημαϊκά μαθήματα ΠΠ

Μελέτη Περιπτώσεων στη Λήψη Αποφάσεων

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Αδειοδότησης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Από τους Μεγάλους Υπολογισμούς στη Λογοτεχνία με Λογισμό Μητρώων

Ε. Γαλλόπουλος

ΤΜΗΥΠ
Πανεπιστήμιο Πατρών

8/5/2015

Στόχοι

Charles Babbage
1791-1871

- **«Μόλις κατασκευαστεί η Αναλυτική Μηχανή, δεν μπορεί παρά να οδηγήσει την εξέλιξη της Επιστήμης. Όποτε της ζητάμε ένα αποτέλεσμα, θα μπαίνει το ερώτημα: Με ποιον τρόπο υπολογισμού με τη μηχανή αυτή θα μπορέσουμε να υπολογίσουμε τα αποτελέσματα στον ελάχιστο χρόνο;» [*Charles Babbage 1864*]**

Ada Byron Lovelace
1815-1852

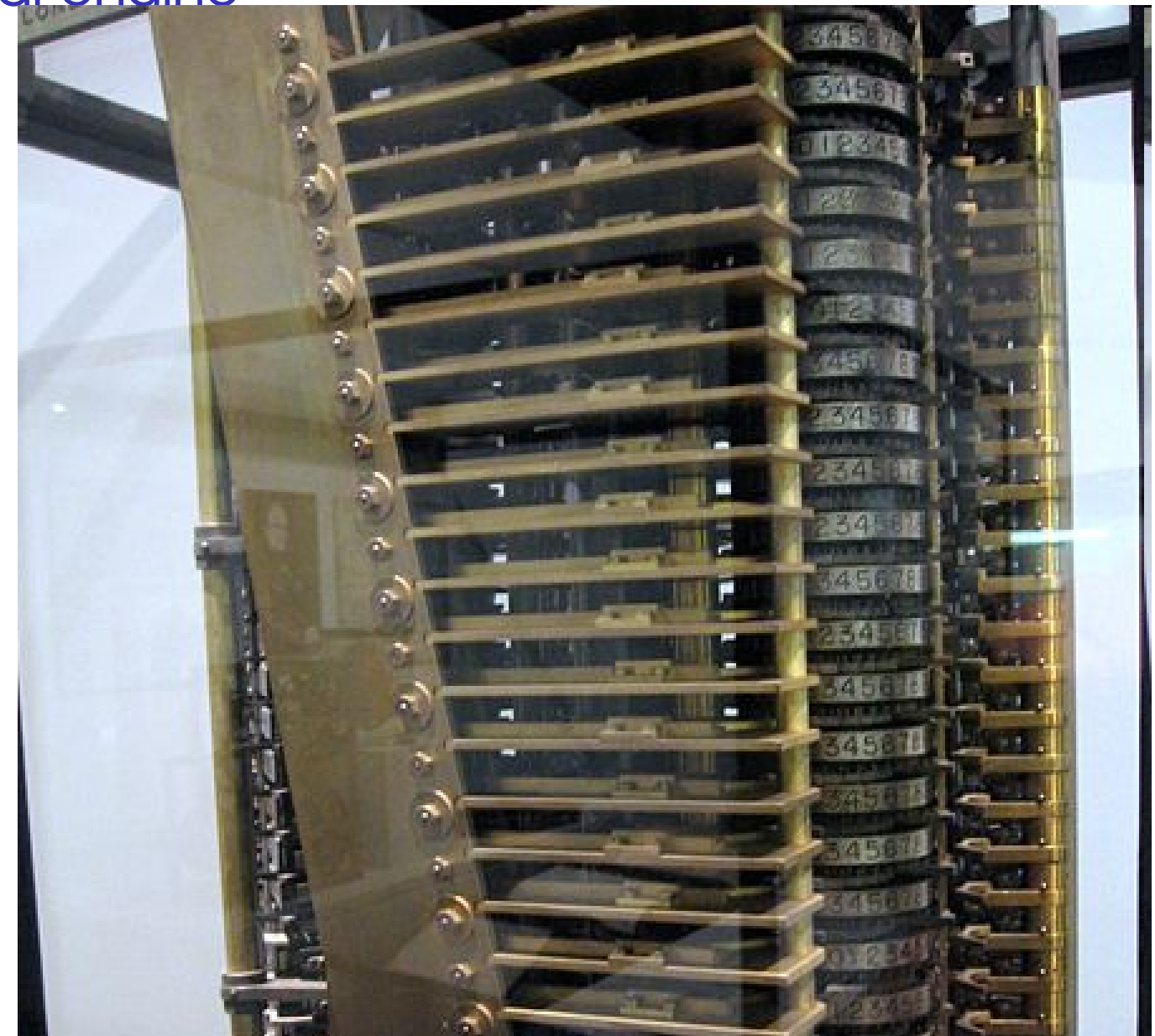
- **Μας ενδιαφέρουν:Ο** σχεδιασμός, **η** ανάπτυξη, **και η** αποδοτική χρήση **υπολογιστικών εργαλείων που βοηθούν στην πρακτική χρήση των** μαθηματικών μοντέλων της επιστήμης και της τεχνολογίας.

- **Να φθάσουμε γρήγορα ... Ταχύτητα**
- **χωρίς να τρακάρουμε ... Ακρίβεια**

Θεμελιώδη προβλήματα φυσικής
Πρόγνωση καιρού
Σχεδιασμός κατασκευών
Σχεδιασμός κυκλωμάτων
Ασφάλεια συστημάτων
Νέα φάρμακα
Γραφικά
Χρηματιστήριο
Παιχνίδια

εξισώσεις – εξισώσεις – εξισώσεις ...

Analytical engine



Top 500 list


www.top500.org
TOP500 SUPERCOMPUTER SITES
panasas
Data + Performance + Private Cloud = Panasas
Intersect360 RESEARCH
DOWNLOAD REPORT
PROJECT
LISTS
STATISTICS
RESOURCES
NEWS
CONTACT
SUBMISSIONS
LINKS
HOME
TOP 10 Systems - 11/2011
1 K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect
2 NUDT YH MPP, Xeon X5670 6C 2.93 GHz, NVIDIA 2050
3 Cray XT5-HE Opteron 6-core 2.6 GHz
4 Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU
5 HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows
6 Cray XE6, Opteron 6136 8C 2.40GHz, Custom
7 SGI Altix ICE 8200EX/8400EX, Xeon HT QC 3.0/Xeon 5570/5670 2.93 Ghz, Infiniband
8 Cray XE6, Opteron 6172 12C 2.10GHz, Custom
9 Bull bullx super-node S6010/S6030
10 BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband
League of TOP 1
Japan's K Computer Tops 10 Petaflop/s to Stay Atop TOP500 List
Fri, 2011-11-11 11:11
BERKELEY, Calif.; KNOXVILLE, Tenn.; and MANNHEIM, Germany (Nov. 14, 2011)—Japan's "K Computer" maintained its position atop the newest edition of the TOP500 List of the world's most powerful supercomputers, thanks to a full build-out that makes it four times as powerful as its nearest competitor. Installed at the RIKEN Advanced Institute for Computational Science (AICS) in Kobe, Japan, the K Computer it achieved an impressive 10.51 Petaflop/s on the Linpack benchmark using 705,024 SPARC64 processing cores.
» Read more
Photos: Awarding of Certificates for No. 1 System in the 38th TOP500 List
Thu, 2011-11-17 19:44
SUPERMICRO
TwinBlade™
20 DP Nodes in 7U Enclosure
intel inside
Xeon
High Performance Computing
FUJITSU
27th ISC'12
June 17 – 21, 2012
Hamburg, Germany
APPRO
Visit Appro Booth #2312 at SC11
LEARN MORE
hp
insideHPC
MEGWARE
customized HPC solutions
Recent Releases
November 2011
June 2011
November 2010

Συχνά τα αποτελέσματα .... προβληματίζουν!

Κουίζ:

$$
\begin{aligned}
10^{20} - 10 - 10^{20} + 20 &= \\
10^{20} + 20 - 10^{20} - 10 &= \\
-10 + 20 - 10^{20} + 10^{20} &= \\
10^{20} - 10^{20} + 20 - 10 &=
\end{aligned}
$$

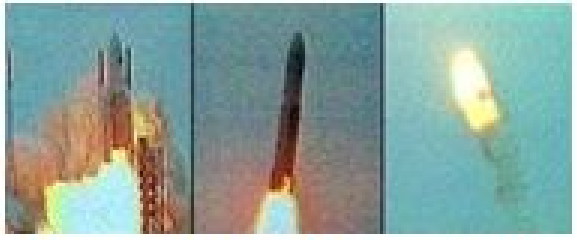
ΠΡΟΒΛΗΜΑΤΙΖΟΥΝ!

Κουίζ:

$$
\begin{aligned}
10^{20} - 10 - 10^{20} + 20 &= 20 \\
10^{20} + 20 - 10^{20} - 10 &= -10 \\
-10 + 20 - 10^{20} + 10^{20} &= 0 \\
10^{20} - 10^{20} + 20 - 10 &= 10
\end{aligned}
$$

Σοβαρές Αστοχίες

Patriot missile failure, 1991
(28 deaths because of bad rounding)



Explosion of Ariane 5, 1996
(500M because of overflow)

Sinking of Sleipner A offshore platform, 1991
(700M because of inaccurate fem approximation)

- Vancouver stock exchange

… and many others

$$$$$ LOSS IN EVALUATING NUMERICAL RELIABILITY

Μεγάλες Εφαρμογές και Υπολογιστικοί Πυρήνες (?)

	1	2	3	4	5	6	7	8	9	
Lattice Gauge (QCD)	*	.	.	*	.	.	.	.	*	.
quantum mechanics	.	.	.	*	.	.	*	*	*	
weather	.	.	.	.	*	*	.	.	.	
CFD	*	.	*	.	*	*	.	.	.	
geodesy	*	*	.	.	.	.	.	.	.	
inverse problems	.	*	.	.	*	.	.	.	.	.
structures	*	.	*	*	.	.	.	.	.	
circuit simulation	*	.	*	.	.	.	*	.	.	
electromagnetics	*	*	*	*	*	*	.	.	.	
financial	*	*	*	.	.	.	*	*	*	
IR	*	*	.	*	.	.	.	.	.	
DS& Image P.	*	*	*	*	*	*	.	.	*	
Internet Algorithmics	*	.	.	*	.	.	.	.	.	

1. γραμμικά συστ.
2. ελάχιστα τετράγωνα
3. μη γραμμικά συστ.
4. ιδιοδιασπάσεις/SVD
5. ταχείς μετασχηματισμοί
6. ταχείς ελλειπτικοί επιλυτές
7. άκαμπτες ΔΕ
8. Μόντε Κάρλο
9. ολοκληρωματικοί μετασχ.

Μητρώα (διπός ρόλος): κωδικοποιούν γραμμικούς μετασχηματισμούς ή πίνακες δεδομένων

Θεωρία, πράξεις και υπολογισμοί με ΜΗΤΡΩΑ

Υπολογιστικοί πυρήνες

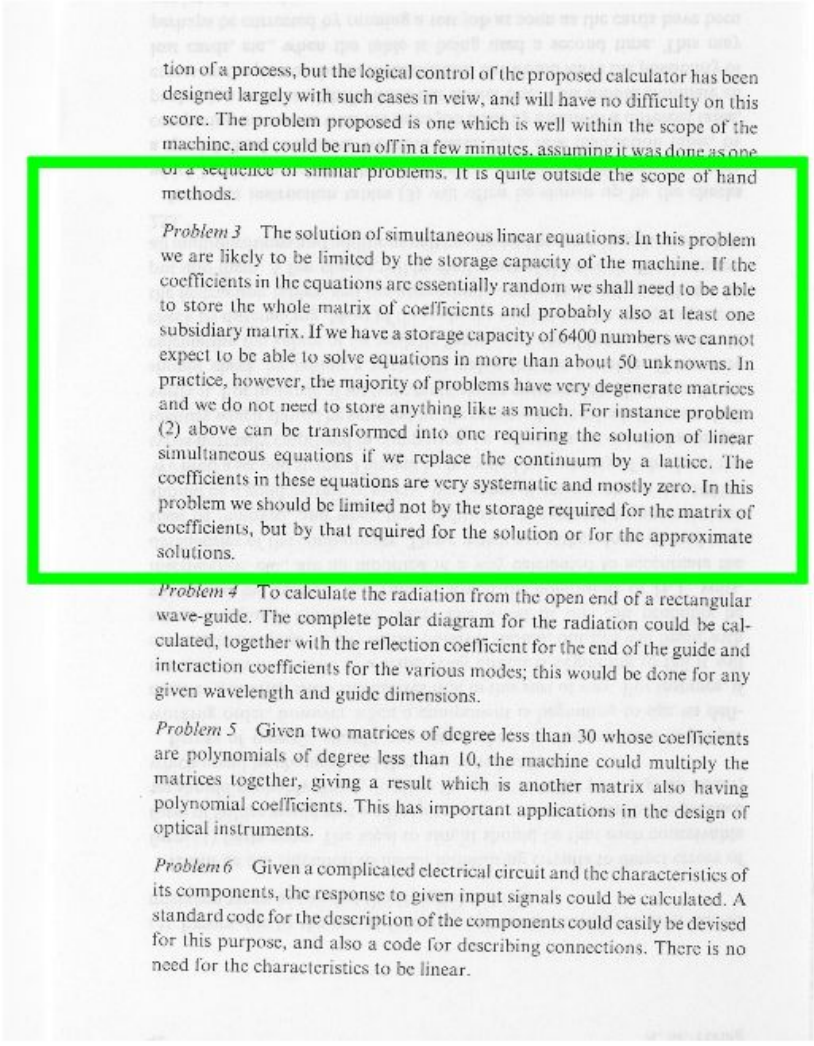
Τμήματα κώδικα που μπορεί να αντιστοιχούν σε κάποια μαθηματική πράξη (π.χ. λύση συστήματος ή FFT) στον οποίο αναλώνεται μεγάλος χρόνος εκτέλεσης σε σχέση με άλλα τμήματα του κώδικα.

Εύρημα

Οι υπολογισμοί με μητρώα είναι υπολογιστικοί πυρήνες των περισσότερων σημαντικών εφαρμογών που απαιτούν υπολογισμούς ευρείας κλίμακος.

γι' αυτό η υπολογιστική Γραμμική Άλγεβρα έχει καθιερωθεί ως το όχημα για την παρουσίαση των τεχνικών του ΕΥ.

Διαπίστωση του Turing



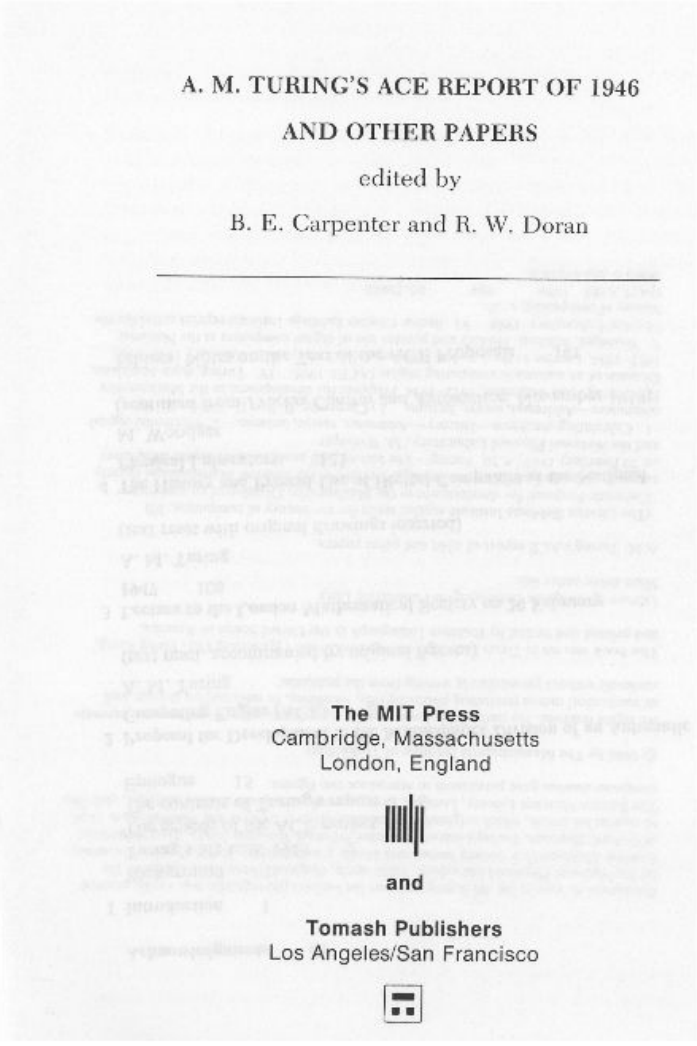
tion of a process, but the logical control of the proposed calculator has been designed largely with such cases in veiw, and will have no difficulty on this score. The problem proposed is one which is well within the scope of the machine, and could be run off in a few minutes, assuming it was done as one of a sequence of similar problems. It is quite outside the scope of hand methods.

Problem 3 The solution of simultaneous linear equations. In this problem we are likely to be limited by the storage capacity of the machine. If the coefficients in the equations are essentially random we shall need to be able to store the whole matrix of coefficients and probably also at least one subsidiary matrix. If we have a storage capacity of 6400 numbers we cannot expect to be able to solve equations in more than about 50 unknowns. In practice, however, the majority of problems have very degenerate matrices and we do not need to store anything like as much. For instance problem (2) above can be transformed into one requiring the solution of linear simultaneous equations if we replace the continuum by a lattice. The coefficients in these equations are very systematic and mostly zero. In this problem we should be limited not by the storage required for the matrix of coefficients, but by that required for the solution or for the approximate solutions.

Problem 4 To calculate the radiation from the open end of a rectangular wave-guide. The complete polar diagram for the radiation could be calculated, together with the reflection coefficient for the end of the guide and interaction coefficients for the various modes; this would be done for any given wavelength and guide dimensions.

Problem 5 Given two matrices of degree less than 30 whose coefficients are polynomials of degree less than 10, the machine could multiply the matrices together, giving a result which is another matrix also having polynomial coefficients. This has important applications in the design of optical instruments.

Problem 6 Given a complicated electrical circuit and the characteristics of its components, the response to given input signals could be calculated. A standard code for the description of the components could easily be devised for this purpose, and also a code for describing connections. There is no need for the characteristics to be linear.



A. M. TURING'S ACE REPORT OF 1946

AND OTHER PAPERS

edited by

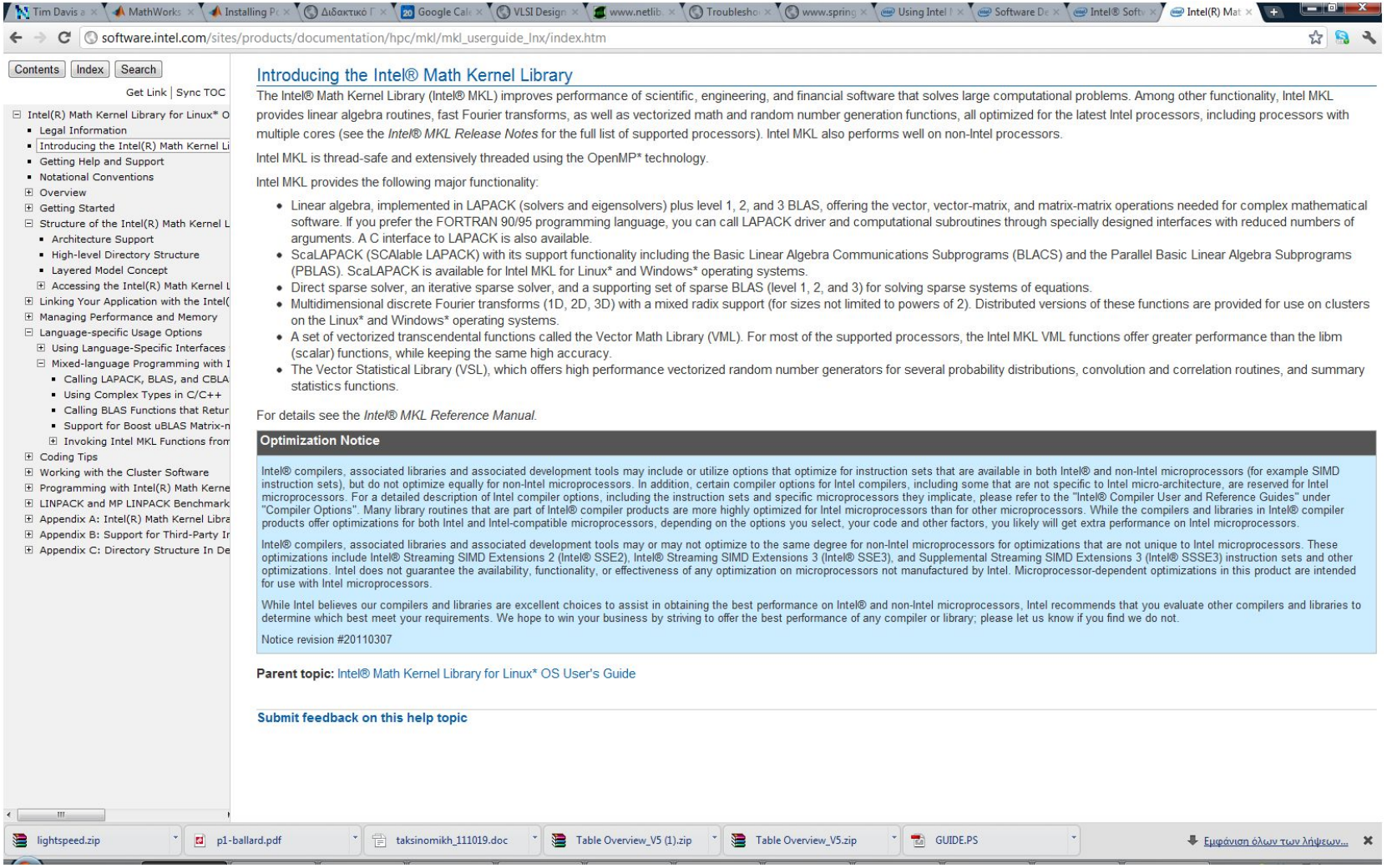
B. E. Carpenter and R. W. Doran

The MIT Press
Cambridge, Massachusetts
London, England

and

Tomash Publishers
Los Angeles/San Francisco

Οι κατασκευαστές αφουγκράζονται



Θεμελιώδη προβλήματα της ΥΓΑ

Βασικές πράξεις γραμμικής άλγεβρας Δίδονται $A \in \mathbb{R}^{n_1 \times n_3}$, $B \in \mathbb{R}^{n_3 \times n_2}$, $C \in \mathbb{R}^{n_1 \times n_2}$, να υπολογιστεί $C + AB$.

Επίλυση γραμμικού συστημάτος Δίδονται $A \in \mathbb{R}^{n \times n}$ και $b \in \mathbb{R}^n$. Να υπολογιστεί $x \in \mathbb{R}^n$ τ.ώ. $Ax = b$.

Γραμμικά ελάχιστα τετράγωνα Δίδονται $A \in \mathbb{R}^{m \times n}$ ανδ $b \in \mathbb{R}^m$. Να υπολογιστεί $x \in \mathbb{R}^n$ που επιλύει $\arg\min_x \|b - Ax\|_2$.

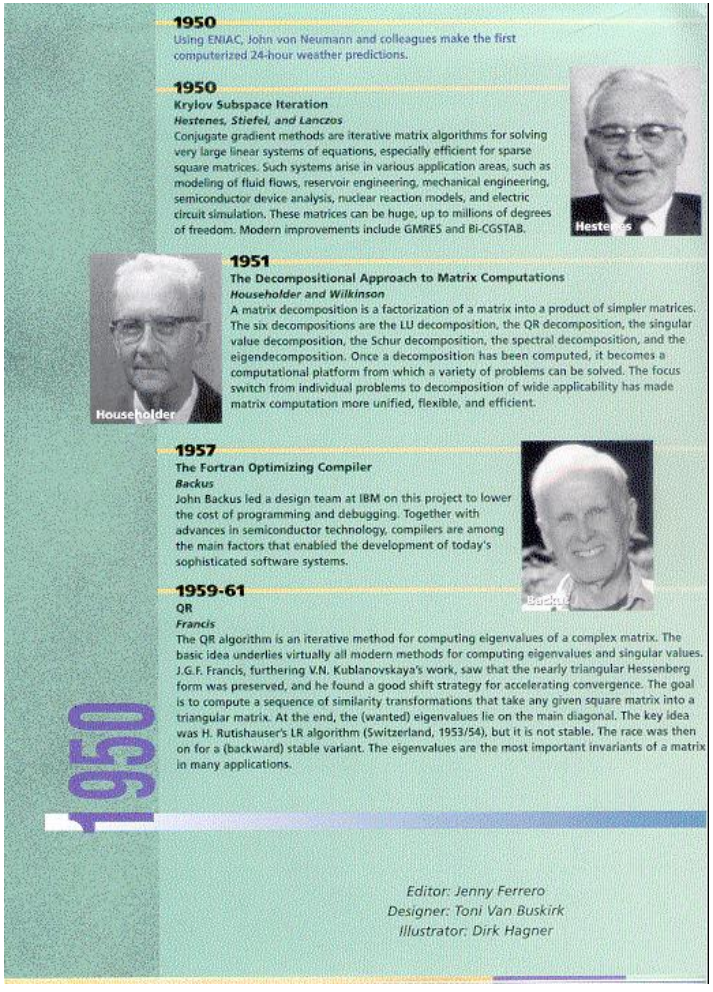
Πρόβλημα ιδιοτιμών Δίδεται $A \in \mathbb{R}^{n \times n}$. Να υπολογιστούν τα ιδιοζεύγη $\lambda \in \mathbb{C}$ και $x \in \mathbb{C}^n$: $Ax = \lambda x$.

Γενικευμένο πρόβλημα ιδιοτιμών Δίδονται $A, B \in \mathbb{R}^{n \times n}$. Να υπολογιστούν $\lambda \in \mathbb{C}$ και $x \in \mathbb{C}^n$: $Ax = \lambda Bx$.

Διάσπαση ιδιαζουσών τιμών Δίδεται $A \in \mathbb{R}^{m \times n}$. Να υπολογιστεί διαγώνιο $\Sigma \in \mathbb{R}^{m \times n}$, ορθογώνια $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$: $A = U\Sigma V^\top$.

Συναρτήσεις μητρώων Δίδονται $A \in \mathbb{R}^{n \times n}$ και συνάρτηση f. Να υπολογιστεί το $f(A)$ ή το $f(A)B$ για $B \in \mathbb{R}^{n \times m}$ ή το $Cf(A)B$ για $C \in \mathbb{R}^{k \times n}$.

Διασπάσεις/παραγοντοποιήσεις μητρώων



1950
Using ENIAC, John von Neumann and colleagues make the first computerized 24-hour weather predictions.

1950
Krylov Subspace Iteration
Hestenes, Stiefel, and Lanczos
Conjugate gradient methods are iterative matrix algorithms for solving very large linear systems of equations, especially efficient for sparse square matrices. Such systems arise in various application areas, such as modeling of fluid flows, reservoir engineering, mechanical engineering, semiconductor device analysis, nuclear reaction models, and electric circuit simulation. These matrices can be huge, up to millions of degrees of freedom. Modern improvements include GMRES and Bi-CGSTAB.

1951
The Decompositional Approach to Matrix Computations
Householder and Wilkinson
A matrix decomposition is a factorization of a matrix into a product of simpler matrices. The six decompositions are the LU decomposition, the QR decomposition, the singular value decomposition, the Schur decomposition, the spectral decomposition, and the eigendecomposition. Once a decomposition has been computed, it becomes a computational platform from which a variety of problems can be solved. The focus switch from individual problems to decomposition of wide applicability has made matrix computation more unified, flexible, and efficient.

1957
The Fortran Optimizing Compiler
Backus
John Backus led a design team at IBM on this project to lower the cost of programming and debugging. Together with advances in semiconductor technology, compilers are among the main factors that enabled the development of today's sophisticated software systems.

1959-61
QR
Francis
The QR algorithm is an iterative method for computing eigenvalues of a complex matrix. The basic idea underlies virtually all modern methods for computing eigenvalues and singular values. J.G.F. Francis, furthering V.N. Kublanovskaya's work, saw that the nearly triangular Hessenberg form was preserved, and he found a good shift strategy for accelerating convergence. The goal is to compute a sequence of similarity transformations that take any given square matrix into a triangular matrix. At the end, the (wanted) eigenvalues lie on the main diagonal. The key idea was H. Rutishauser's LR algorithm (Switzerland, 1953/54), but it is not stable. The race was then on for a (backward) stable variant. The eigenvalues are the most important invariants of a matrix in many applications.

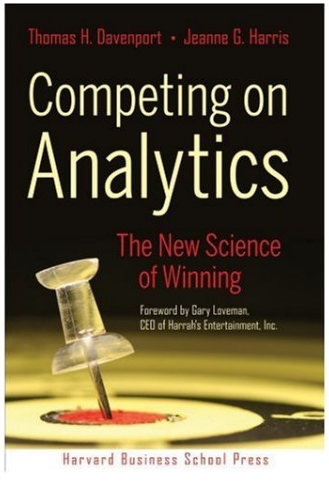
1950

Editor: Jenny Ferrero
Designer: Toni Van Buskirk
Illustrator: Dirk Hagner

Οι μεγάλες 6 ... και μερικές ακόμα

LU, PLU	
$LL^\top$	Cholesky
QR, QRP	
$V\Lambda V^\top$, $V\Lambda V^{-1}$	φασματική
$VTV^\top$	Schur
$U\Sigma V^\top$	SVD
QM	πολική
QU	μη αρνητική

Επιχειρηματικό Κίνητρο



analytics εννοούνται η εκτενής χρήση δεδομένων, στατιστικών τεχνικών και ποσοτικής ανάλυσης, ερμηνευτικών και προβλεπτικών μοντέλων και διαχείριση βασισμένη στα δεδομένα για τη λήψη αποφάσεων ...

Αναλυτικές τεχνικές: στατιστικοί και ποσοτοτικοί αλγόριθμοι, εξόρυξης δεδομένων, ανάκτηση κειμενικής πληροφορίας, κατηγοριοποίηση κειμένων, ...

Το βραβείο των $1.000.000

With Netflix you can rent as many DVDs as you want and watch movies instantly on your PC for one low price (...) no late fees (...) no due dates, and DVD shipping free both ways. Plans start from only $4.99 (...) With our most popular plan, you can rent as many DVDs as you want (3 at-a-time) (...) all for just $17.99 a month +tax. ...

Terms and Conditions in a Nutshell

- Contest begins October 2, 2006 and continues through at least October 2, 2011.
- Contest is open to anyone, anywhere (except certain countries listed below).
- You have to register to enter.
- Once you register and agree to these Rules, you'll have access to the Contest training data and qualifying test sets.
- To qualify for the $1,000,000 Grand Prize, the accuracy of your submitted predictions on the qualifying set must be at least 10% better than the accuracy Cinematch can achieve on the same training data set at the start of the Contest.
- To qualify for a year's $50,000 Progress Prize the accuracy of any of your submitted predictions that year must be less than or equal to the accuracy value established by the judges the preceding year.
- To win and take home either prize, your qualifying submissions must have the largest accuracy improvement verified by the Contest judges, you must share your method with (and non-exclusively license it to) Netflix, and you must describe to the world how you did it and why it works.

Λύση

The BigChaos Solution to the Netflix

Andreas Töscher and Michael Jahrer

commendo research & consulting
Neuer Weg 23, A-8580 Köflach, Aust
{andreas.toescher,michael.jahrer}@comme

Input: A matrix $\mathbf{P}$ with all previous probe predictions. $\mathbf{P}$ always includes a constant p
column with ones).

Exclude the Probe ratings $\mathbf{r}$ from the training set.
Initialize the weights.
$RMSE_{best} = \infty$
$RMSE_{epoch} = 1000$
$epochs = 0$
while $RMSE_{epoch} \leq RMSE_{best}$ **do**
  Train one epoch.
  if $RMSE_{epoch} \leq RMSE_{best}$ **then**
    $RMSE_{best} = RMSE_{epoch}$
    Save the current weights.
  end
  Predict the probe set $\hat{\mathbf{p}}$.
  Merge current probe prediction $\hat{\mathbf{p}}$ and previous predictions: $\mathbf{X} = [\mathbf{P}\ \hat{\mathbf{p}}]$
  Calculate blending weights: $\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{r}$
  Calculate prediction of the current blend: $\mathbf{p} = \mathbf{X} \cdot \mathbf{w}$
  Calculate the RMSE of the blend: $RMSE_{epoch} = \sqrt{\frac{1}{R}\sum_{i=1}^{R}(p_i - r_i)^2}$; r_i is probe r
  is #ratings in the probe set
  $epochs = epochs + 1$

Το κίνητρο των $25.000.000.000

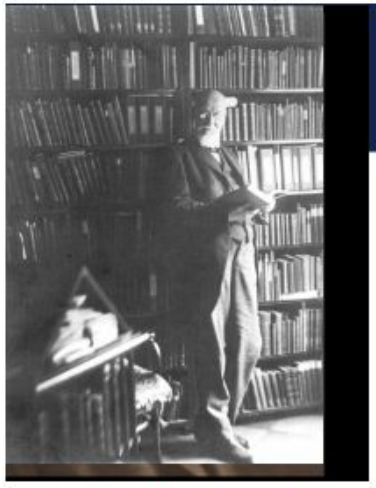
THE $25,000,000,000* EIGENVECTOR
THE LINEAR ALGEBRA BEHIND GOOGLE

KURT BRYAN† AND TANYA LEISE‡

Abstract. Google's success derives in large part from its PageRank algorithm, which ranks the importance of webpages according to an eigenvector of a weighted link matrix. Analysis of the PageRank formula provides a wonderful applied topic for a linear algebra course. Instructors may assign this article as a project to more advanced students, or spend one or two lectures presenting the material with assigned homework from the exercises. This material also complements the discussion of Markov chains in matrix algebra. Maple and Mathematica files supporting this material can be found at www.rose-hulman.edu/~bryan.

Key words. linear algebra, PageRank, eigenvector, stochastic matrix

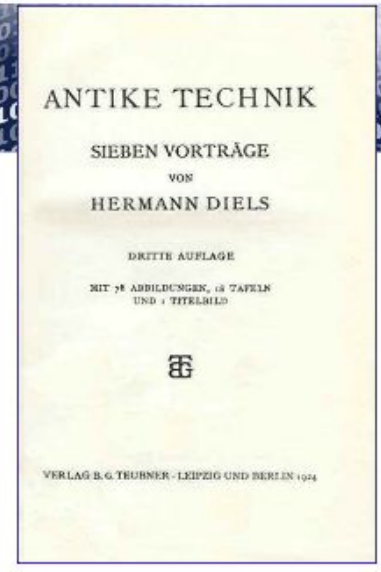
AMS subject classifications. 15-01, 15A18, 15A51



. Diels, 1848-1922

- «Αλλά ακόμα και να υποθέταμε ότι ήταν διαθέσιμες ... όλες οι αρχαίες εκδόσεις από τον Όμηρο (8ο π.Χ.) ως το Νόννο (5 μ.Χ.) .... και ότι ένα γιγαντιαίο προσωπικό από μελετητές τις είχε επεξεργαστεί και αποδελτιώσει και .... ότι είχαν αποθηκευτεί κάπου σε χιλιάδες κουτιά, από πού θα προέκυπταν ο χρόνος, τα χρήματα και η δύναμη για να εξεταστούν σχολαστικά τόσα εκατομμύρια δελτία ώστε να υπάρξει «Νούς» στο Χάος;

[Hermann Diels, *1905*]



Θησαυρός της Ελληνικής Γλώσσας
(TLG: Thesaurus Lingua Graecae)

- Συλλογή όλων των ελληνικών κειμένων (8ο π.Χ – 6ο μ.Χ.) και των σχολίων, ιστοριογραφικών και λεξικογραφικών έργων μέχρι το 1453.
- Άνω των 12.000 έργων 3.700 συγγραφέων, περισσότερες από 75Μ λέξεις

Πώς αντλούμε πληροφορίες από μια βιβλιοθήκη;

- Με ποιόν τρόπο;
- Πόσο ακριβείς είναι οι πληροφορίες;
- Πόσο γρήγορα;
- Πώς μοντελοποιούμε την πληροφορία;

Τυπικά βήματα στην Ανάκτηση Πληροφορίας

(i) προετοιμασία κειμένου
→ κατασκευή `μητρώου όρων-κειμένων' (MOK ή tdm)

Τυπικά βήματα στην Ανάκτηση Πληροφορίας

(I) προετοιμασία κειμένου
→ κατασκευή `μητρώου όρων-κειμένων' (MOK ή tdm)

(II) εφαρμογή αλγορίθμου Ανάκτηση Πληροφορίας
→ ομαδοποίηση ή απάντηση σε ερώτηση

Μοντέλο Διανυσματικού Χώρου (VSM: (Salton'68))

- Κείμενα αναπαριστώνται ως διανύσματα με όρους (συναρτήσεις) των συχνοτήτων εμφάνισής τους
- Συλλογές κειμένων αναπαριστώνται ως `μητρώα όρων-κειμένων' (μοκ).
- Για την ανάκτηση πληροφορίας, κατασκευάζουμε το διάνυσμα ερώτησης q (ψευδο-κείμενο)
- Στόχος: Εύρεση κειμένων πλησιέστερων στο q.
- Μετρική: Γωνία μεταξύ q και διανύσματος κειμένου d: $\frac{q^\top d}{\|q\|_2 \|d\|_2}$.

Μοντέλο Διανυσματικού Χώρου (VSM: (Salton'68))

- Κείμενα αναπαριστώνται ως διανύσματα με όρους (συναρτήσεις) των συχνοτήτων εμφάνισής τους
- Συλλογές κειμένων αναπαριστώνται ως `μητρώα όρων-κειμένων' (μοκ).
- Για την ανάκτηση πληροφορίας, κατασκευάζουμε το διάνυσμα ερώτησης q (ψευδο-κείμενο)
- Στόχος: Εύρεση κειμένων πλησιέστερων στο q.
- Μετρική: Γωνία μεταξύ q και διανύσματος κειμένου d: $\frac{q^\top d}{\|q\|_2 \|d\|_2}$.

TERM	d_1	d_2
and	1	0
as	0	1
beyond	1	0
documents	0	1
frequencies	0	1
model	1	0
of	0	1
represented	0	1
salton	1	0
space	1	0
vector	1	0
vectors	0	1
vsm	1	0
weighted	0	1
word	0	1
$A \in \mathbb{R}^{15\times 2}, \mathrm{nnz}(A) = 15$		

d_1 Vector Space Model VSM: ((Salton'68) and beyond)

d_2 Documents represented as vectors of weighted word frequencies.

TERM	d_1	d_2
and	1	0
as	0	1
beyond	1	0
documents	0	1
frequencies	0	1
model	1	0
of	0	1
represented	0	1
salton	1	0
space	1	0
vector	1	0
vectors	0	1
vsm	1	0
weighted	0	1
word	0	1

$A \in \mathbb{R}^{15\times 2}, \mathrm{nnz}(A) = 15$

αφαίρεση
$\Rightarrow$
stopwords

TERM	d_1	d_2
documents	0	1
frequencies	0	1
model	1	0
represented	0	1
salton	1	0
space	1	0
vector	1	0
vectors	0	1
vsm	1	0
weighted	0	1
word	0	1

$A \in \mathbb{R}^{11\times 2}, \mathrm{nnz}(A) = 11$

TERM	d_1	d_2
documents	0	1
frequencies	0	1
model	1	0
represented	0	1
salton	1	0
space	1	0
vector	1	0
vectors	0	1
vsm	1	0
weighted	0	1
word	0	1
$A \in \mathbb{R}^{11\times 2}, \mathrm{nnz}(A) = 11$		

TERM	d_1	d_2
documents	0	1
frequencies	0	1
model	1	0
represented	0	1
salton	1	0
space	1	0
vector	1	0
vectors	0	1
vsm	1	0
weighted	0	1
word	0	1

$A \in \mathbb{R}^{11\times 2}, \mathrm{nnz}(A) = 11$

εφαρμογή
$\Rightarrow$
stemming

TERM	d_1	d_2
document	0	1
frequenc	0	1
model	1	0
repres	0	1
salton	1	0
space	1	0
vector	1	1
vsm	1	0
weight	0	1
word	0	1

$A \in \mathbb{R}^{10\times 2}, \mathrm{nnz}(A) = 11$

TERM	d_1	d_2
document	0	1
frequenc	0	1
model	1	0
repres	0	1
salton	1	0
space	1	0
vector	1	1
vsm	1	0
weight	0	1
word	0	1
$A \in \mathbb{R}^{10\times 2}, \mathrm{nnz}(A) = 11$		

προσθήκη
$\Rightarrow$
3ου κειμ.

TERM	d_1	d_2	d_3
collect	0	0	1
document	0	1	2
frequenc	0	1	0
matric	0	0	1
model	1	0	0
repres	0	1	1
salton	1	0	0
space	1	0	0
tdm	0	0	1
term	0	0	1
vector	1	1	0
vsm	1	0	0
weight	0	1	0
word	0	1	0
$A \in \mathbb{R}^{14\times 3}, \mathrm{nnz}(A) = 17$			

Παρατηρήσεις

- Ένα ΜΟΚ για λεξικό m όρων και n κειμένων παριστάται με μητρώο $A \in \mathbb{R}^{m \times n}$.
- Η Ανάκτηση Πληροφορίας χρησιμοποιεί εντατικά τα εργαλεία της Γραμμικής Άλγεβρας

Ειδικά χαρακτηριστικά των ΜΟΚ

- Μεγάλο μέγεθος και μεγάλη αραιότητα (περί τα 90% των στοιχείων μηδενικά)
- Μη αρνητικά μητρώα

Προσέγγιση Μητρώων

Μορφή γινομένου μητρώων τάξης r

Κάθε μητρώο $A \in \mathbb{R}^{m \times n}$ τάξης $r \leq \min\{m, n\}$ μπορεί να εκφραστεί ως $A = BC$ όπου $B \in \mathbb{R}^{m \times r}, C \in \mathbb{R}^{r \times n}$. Οι παράγοντες B, C δεν είναι μοναδικοί. Ο παράγοντας B (C) δεν μπορεί να έχει λιγότερες από r στήλες (γραμμές).

Κάθε μητρώο A τάξης r του οποίου γνωρίζουμε την Αναγμένων Γραμμών Κλιμακωτή Μορφή R μπορεί να γραφτεί ως γινόμενο δύο μητρώων $m \times r$ επί $r \times n$:

A = (r στήλες οδηγών του A) $\times$ (πρώτες r γραμμές του R).

Παράδειγμα

Πριν είδαμε ότι

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \Rightarrow R = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

είναι το ΑΓΚΜ του A, οπότε $r = 2$ και ότι οι στήλες οδηγοί είναι οι $1, 3$.
Διαπιστώνουμε ότι

$$A = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \end{pmatrix}$$

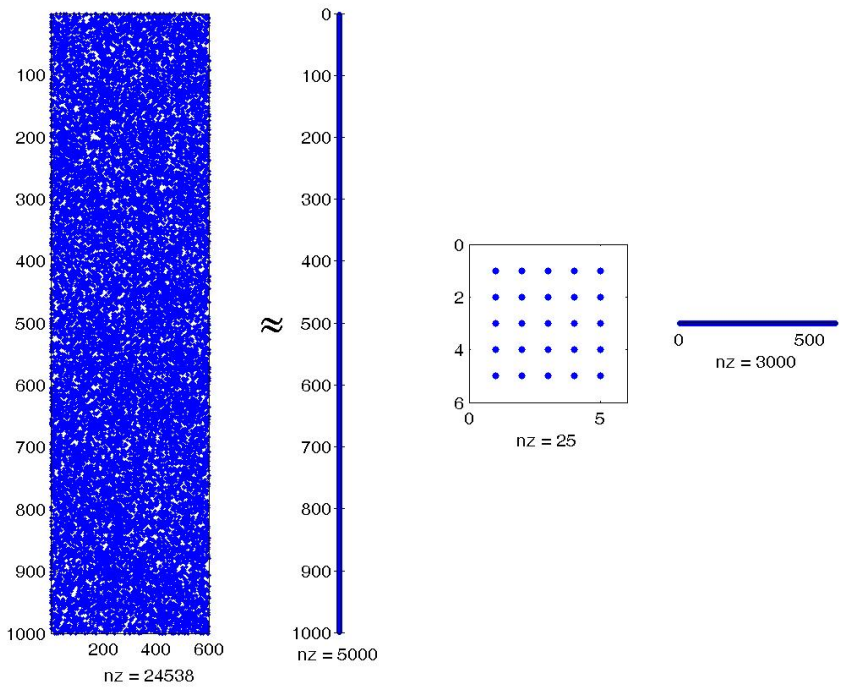
ΣΗΜΑΝΤΙΚΟ: Ένα μητρώο μικρής τάξης μπορεί να γραφτεί ως γινόμενο μητρώων μειωμένης διάστασης

Πρόβλημα Προσέγγισης Μητρώων

Κίνητρο: Τα μητρώα είναι τεράστια

Στόχος: Φθηνές προσεγγίσεις;

Εργαλείο: Παραγοντοποιήσεις μικρής τάξης: $A \approx XWZ$



Διάσπαση ιδιάζουσας τιμής (SVD)

Θεώρημα

Κάθε μητρώο $A \in \mathbb{R}^{m \times n}$ μπορεί να παραγοντοποιηθεί ως γινόμενο 2 ορθογώνιων U, V και ενός μη αρνητικού διαγώνιου μητρώου Σ μητρώων:

$$A = U \begin{pmatrix} \hat{\Sigma} \\ 0 \end{pmatrix} V^\top,$$

όπου $U^\top U = I_m, V^\top V = I_n, \hat{\Sigma} = \mathrm{diag}[\sigma_1, \ldots, \sigma_n] \geq 0.$

Από το SVD μαθαίνουμε σχεδόν τα πάντα για το μητρώο

Ιδιότητες

Από τον ορισμό

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^\top$$

«τηλεσκοπικά»:

$$\begin{aligned}
A &\approx A^{(1)} := \sigma_1 u_1 v_1^\top \\
A &\approx A^{(2)} := A^{(1)} + \sigma_2 u_2 v_2^\top &= U_2 \Sigma_2 V_2^\top \\
&\vdots \\
A &\approx A^{(s)} := A^{(s-1)} + \sigma_s u_s v_s^\top &= U_s \Sigma_s V_s^\top \\
&\vdots \\
A &= A^{(r)} := A^{(r-1)} + \sigma_r u_r v_r^\top &= U \Sigma V^\top
\end{aligned}$$

Προσέγγιση μέσω SVD (Schmidt'07, EckartYoung'36)

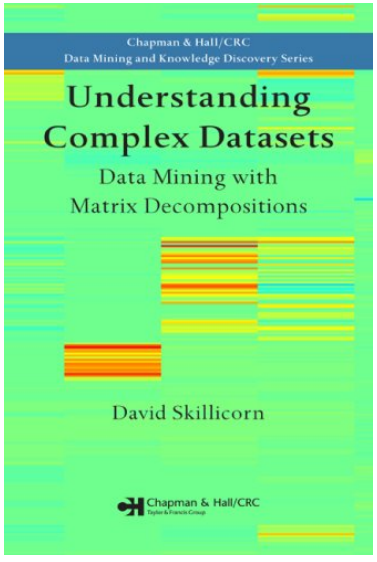
Θεώρημα

$$U_k \Sigma_k V_k^\top = \arg \min_{\mathrm{rank}(X)=k} \|A - X\|$$

με σφάλμα προσέγγισης

$$\|A - U_k \Sigma_k V_k^\top\|_2 = \sigma_{k+1}$$

$$\|A - U_k \Sigma_k V_k^\top\|_F = \left(\sum_{j=k+1}^{r} \sigma_j^2 \right)^{1/2}$$



Μη αρνητικά μητρώα

Θεωρία Perron-Frobenius για το φάσμα

Θέμα: Στις εφαρμογές τα μητρώα συχνά είναι μη αρνητικά ή αυστηρά θετικά. Τότε διαθέτουν πολλές απροσδόκητες και χρήσιμες ιδιότητες.

Μη αρνητικά μητρώα

Θεωρία Perron-Frobenius για το φάσμα

Θέμα: Στις εφαρμογές τα μητρώα συχνά είναι μη αρνητικά ή αυστηρά θετικά. Τότε διαθέτουν πολλές απροσδόκητες και χρήσιμες ιδιότητες.

Κάθε $A > 0 \in \mathbb{R}^{n \times n}$ έχει τουλάχιστον μία μηδενική ιδιοτιμή.

Απόδειξη Έστω $A > 0$. Τότε

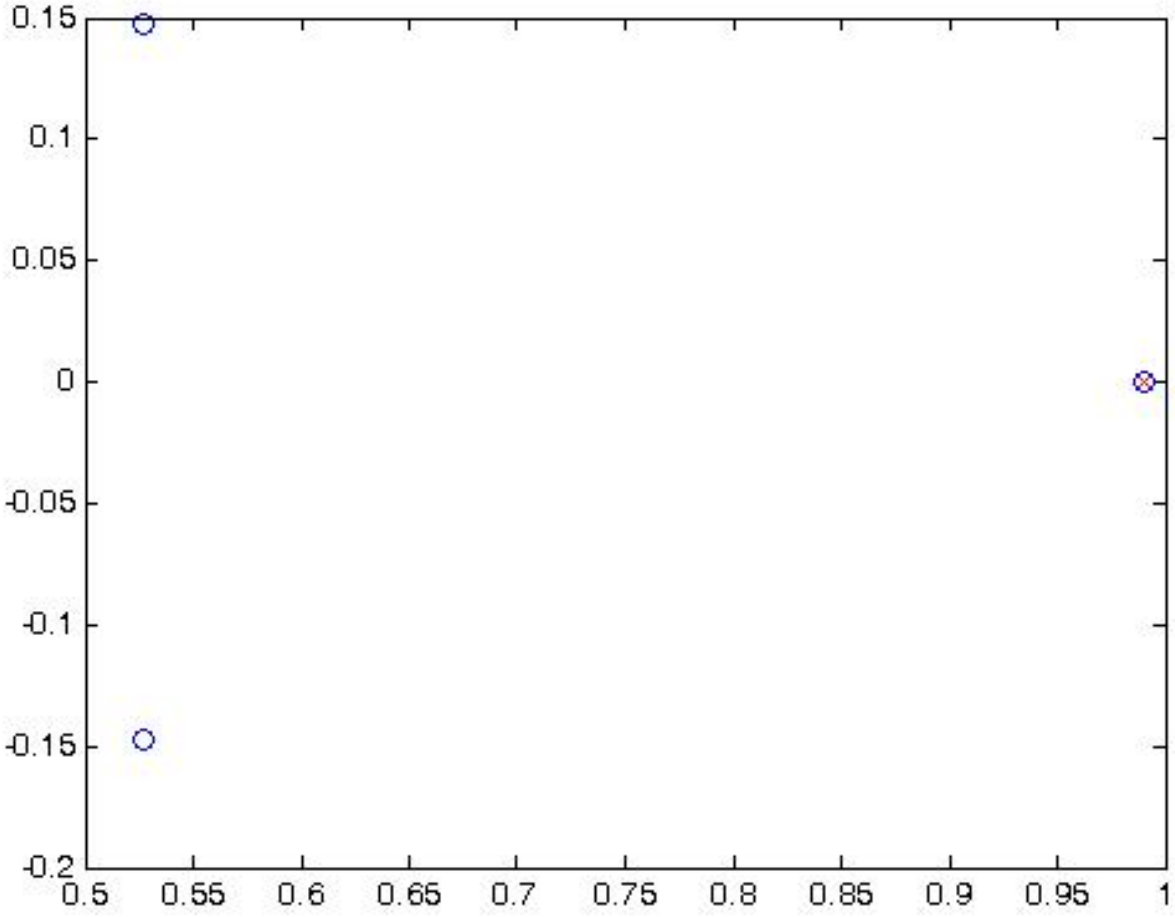
- Αν όλες οι ιδιοτιμές ήταν 0
- ... η κανονική μορφή Jordan του $X^{-1}AX = J$ θα είχε μηδενική διαγώνιο,
- ... οπότε οπωσδήποτε $A^n = 0$, που είναι αδύνατο καθώς $A > 0$.

O. Perron, In Zur Theorie der Matrizen, Math. Ann. 64 (1907), 248-263.

Πείραμα

Μητρώα με τυχαία στοιχεία από $U(0,1)$ (MATLAB `rand(n)`)

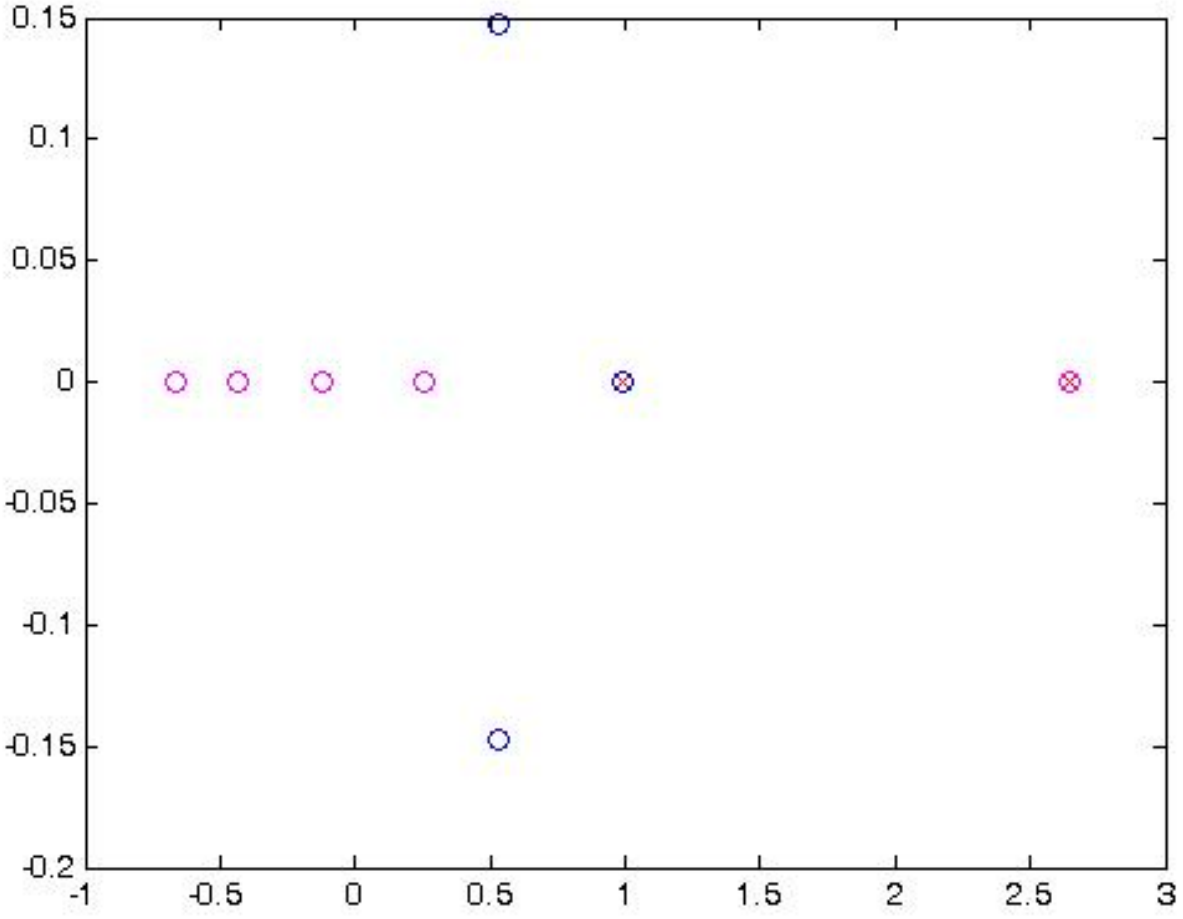
$n = 3$



Πείραμα

Μητρώα με τυχαία στοιχεία από $U(0,1)$ (MATLAB `rand(n)`)

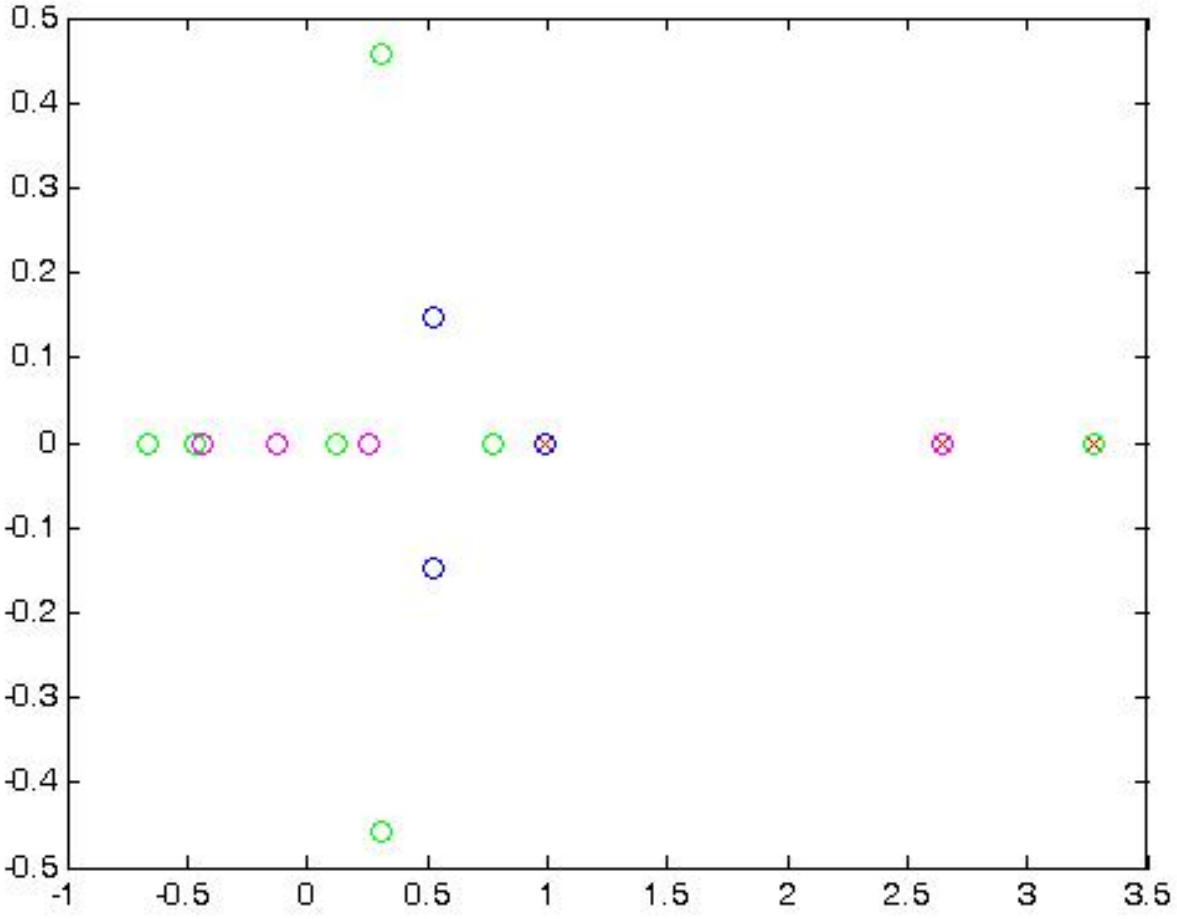
$n = 3, 5$



Πείραμα

Μητρώα με τυχαία στοιχεία από $U(0,1)$ (MATLAB `rand(n)`)

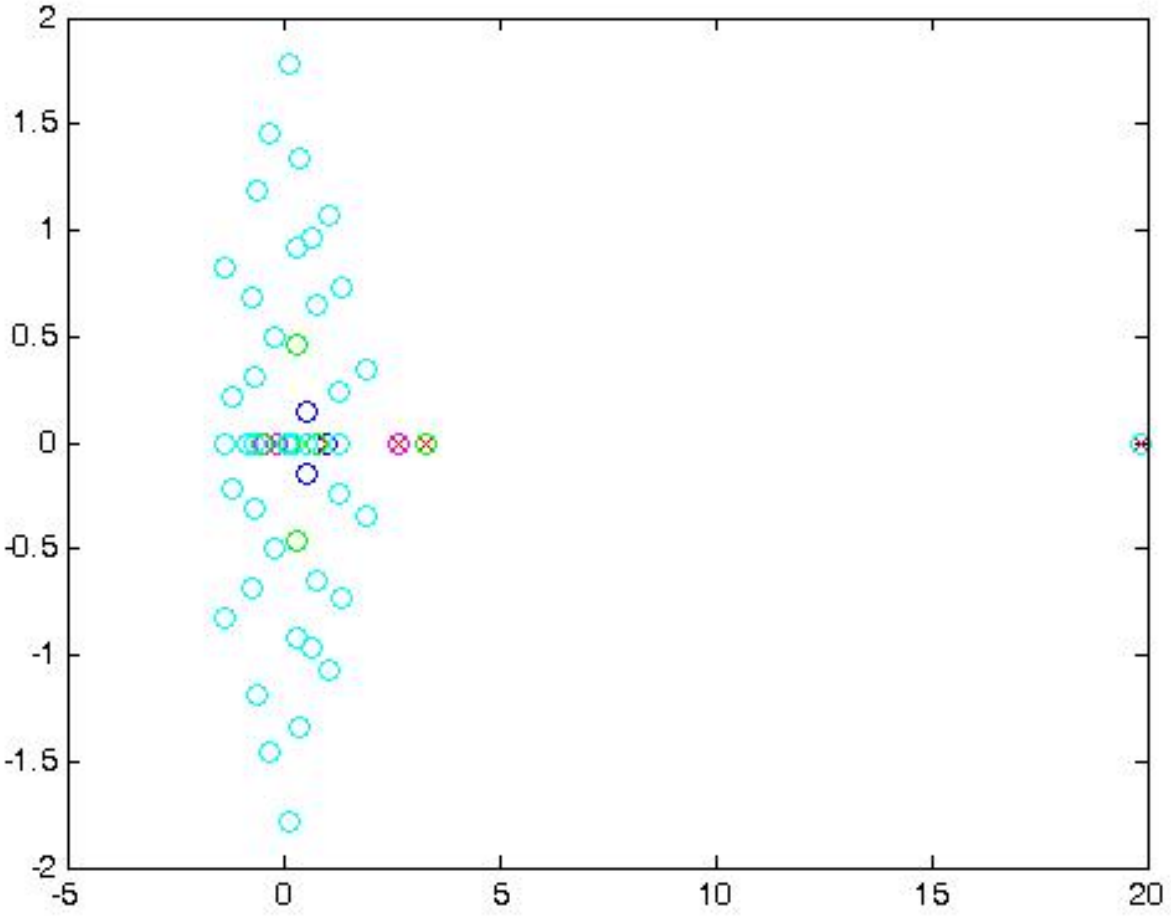
$n = 3, 5, 7$



Πείραμα

Μητρώα με τυχαία στοιχεία από $U(0,1)$ (MATLAB `rand(n)`)

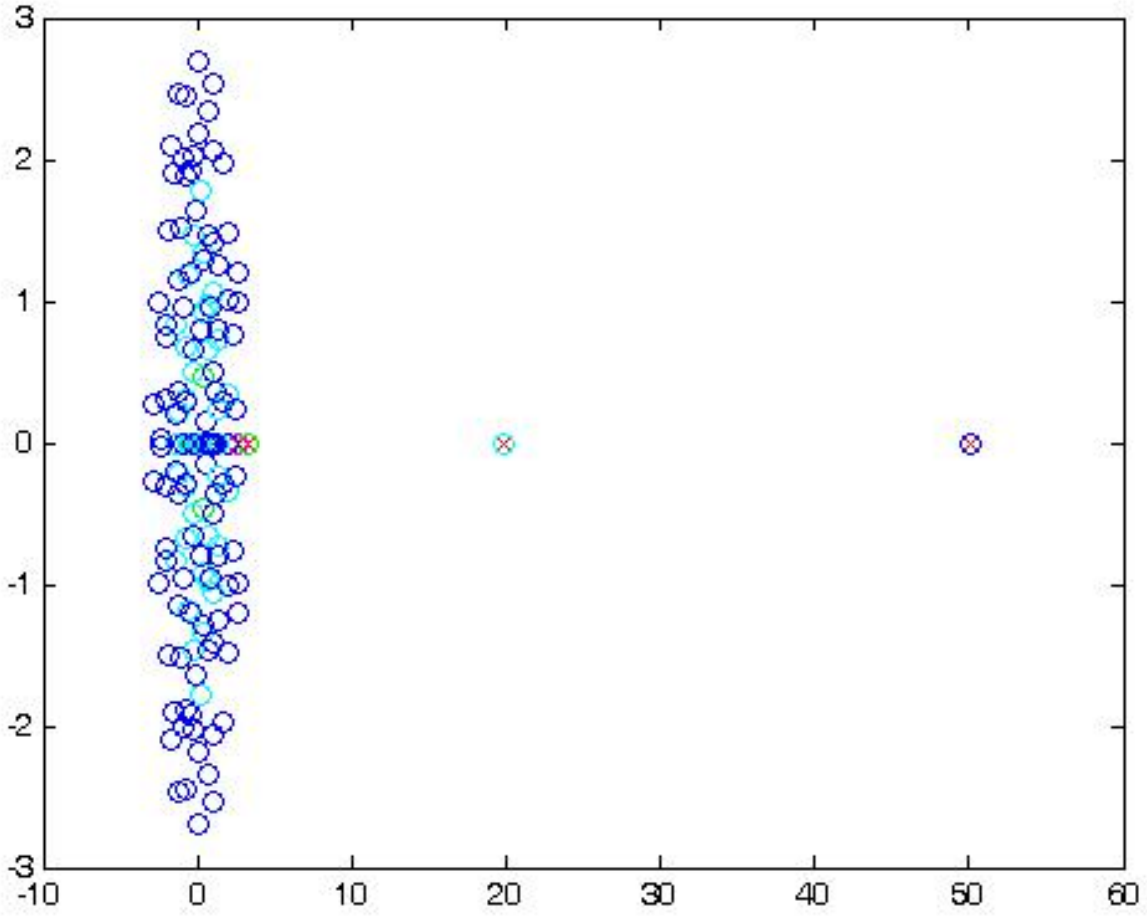
$n = 3, 5, 7, 40$



Πείραμα

Μητρώα με τυχαία στοιχεία από $U(0,1)$ (MATLAB `rand(n)`)

$n = 3, 5, 7, 40, 100$



Θεώρημα Perron για θετικά μητρώα (Perron'1905)

Μερικά αποτελέσματα

Αν $A > 0$ και $r = \rho(A)$ (φασματική ακτίνα) τότε:

1. $r > 0$
2. $r \in \lambda(A)$
3. Η αλγεβρική πολλαπλότητα του $\rho(A)$ είναι 1.
4. Υπάρχει ιδιοδιάνυσμα $x > 0$ τ.ώ. $Ax = rx$. Αυτό είναι το μοναδικό θετικό ιδιοδιάνυσμα.
5. Το διάνυσμα Perron p είναι το μοναδικό διάνυσμα για το οποίο ισχύει ότι

$$Ap = rp, p > 0, \text{ και } \|p\|_1 = 1.$$

6. r είναι η μοναδική ιδιοτιμή του A με μέτρο $|r| = 1$.

Μη αρνητικά μητρώα

Πρόκληση: Να επεκταθούν οι παραπάνω ιδιότητες σε μη αρνητικά μητρώα;

Μη αρνητικά μητρώα

Πρόκληση: Να επεκταθούν οι παραπάνω ιδιότητες σε μη αρνητικά μητρώα;
Μη προφανές! Αν

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

τότε

- $r = 0$ (παραβαίνει την 1η ιδιότητα Perron)
- η αλγεβρική πολλαπλότητα είναι 2 (παραβαίνει την 3η ιδιότητα Perron)
- $x = [1, 0]^\top$ είναι το μοναδικό ιδιοδιάνυσμα για το οποίο $e^\top x = 1$, αλλά το x δεν είναι θετικό (παραβαίνει την 4η ιδιότητα Perron)

Αν

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

τότε $\lambda(A) = \pm i$ άρα υπάρχουν 2 ιδιοτιμές ίσες με 1 σε απόλυτη τιμή (παραβαίνει την 6η ιδιότητα Perron).

Η μάχη δεν χάθηκε

Χωρίς περαιτέρω υποθέσεις μπορούμε να δείξουμε ότι

Αν $A \geq 0 \in \mathbb{R}^{n \times n}$ και $r = \rho(A)$ τότε $r \in \lambda(A)$ και υπάρχει αντίστοιχο ιδιοδιάνυσμα $x \geq 0$ τέτοιο ώστε $Ax = rx$.

Ο Frobenius συνειδητοποίησε ότι τα προβλήματα οφείλονταν όχι μόνον στην ύπαρξη μηδενικών, αλλά στη θέση αυτών μέσα στο μητρώο (G. Frobenius, em Ueber Matrizen aus nicht negativen Elementen, S.-B. Preuss Acad. Wiss. Berlin (1912), 456-477.)

Αναγωγήσιμα μητρώα

Ορισμός

- Ένα μητρώο A καλείται αναγωγήσιμο αν υπάρχει μεταθετικό μητρώο P ώστε το $P^\top AP$ να είναι κατά πλοκάδες άνω τριγωνικό, διαφορετικά καλείται μη αναγωγήσιμο.
- δηλ. ανν

$$P^\top AP = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

- Αν ένα μητρώο είναι μη αναγωγήσιμο, κάθε γραμμή και κάθε στήλη θα έχει τουλάχιστον ένα μη μηδενικό στοιχείο πέραν της διαγωίου.
- $A \in \mathbb{R}^{n\times n} \geq 0$ είναι μη αναγωγήσιμο ανν $(I+A)^{n-1} > 0$.
- Υπάρχουν αλγόριθμοι κόστους αναγωγής μητρώου σε σε κατά πλοκάδες άνω τριγωνική μορφή (Tarjan, βασισμένες σε γραφοθωρία με DFS κόστους $O(n+nnz)$
- Αναγωγησιμότητα $\Rightarrow$ αναγωγή ορισμένων προβλημάτων σε μικρότερα αλλά περισσότερα.

Επέκταση φασματικών ιδιοτήτων

(Frobenius'12)

Θεώρημα Perron-Frobenius

Αν $A \geq 0$ και μη αναγωγήσιμο ισχύουν τα παρακάτω:

1. $r = \rho(A) \in \lambda(A)$ και $r > 0$.
2. Η αλγεβρική πολλαπλότητα του $\rho(A)$ είναι 1.
3. Υπάρχει ιδιοδιάνυσμα $x > 0$ τέτοιο ώστε $Ax = rx$.
4. Το διάνυσμα Perron είναι το μοναδικό διάνυσμα p που ικανοποιεί

$$Ap = rp, p > 0, \text{ και } \|p\|_1 = 1.$$

   Δεν υπάρχουν άλλα μη αρνητικά ιδιοδιανύσματα του A εκτός από θετικά πολλαπλάσια του p.
5. Το $\rho(A)$ αυξάνει αν αυξήσουμε οποιοδήποτε στοιχείο του A.

Στοχαστικά μητρώα

Κίνητρο: Πολλές εφαρμογές οδηγούν σε μητρώα με στοιχεία που είναι πιθανότητες. Αυτά είναι θετικά και στοχαστικά (κατά στήλες ή κατά γραμμές).

Ορολογία: Ένα διάνυσμα ή μητρώο, $A \geq 0$, καλείται στοχαστικό κατά γραμμές όταν το άθροισμα των στοιχείων κάθε γραμμής ισούται με 1.

Παραδείγματα: Στοχαστικές διαδικασίες, ουρές Markov

Υπενθύμιση: Perron-Frobenius Αν $A \in \mathbb{R}^{n \times n}$ είναι στοχαστικό τότε $\rho(A) = 1 = \lambda_{max}$ όπου το λ_{max} αποκαλείται ρίζα Perron και ικανοποιεί $Ap = \lambda_{max}(A)p$ όπου $p > 0$ είναι στοχαστικό.

Εργοδικό θεώρημα

Αν ένα μητρώο είναι μη αναγωγήσιμο, στοχαστικό κατά στήλες (δηλ $e^\top A = e^\top$) και η μοναδική ιδιοτιμή $\lambda_{max} = 1$ είναι μεγαλύτερη όλων των άλλων σε μέτρο, τότε

$$\lim_{k \to \infty} A^k = pe^\top$$

Analysis of Networks: Motivation

Importance:

Networks can be used to describe and analyze interactions between various objects (nodes). Notable examples include applications in:

- sociology
- biology
- World Wide Web
- Communications

Thus...

...analysis of networks (especially of the complex ones) becomes crucial for scientists and practicioners.

Important characteristics?

Among others: a) Relative importance of a node within the network (centrality), b) Easiness of "flow" among the nodes (communicability) c) special relations among the nodes (i.e. a cluster)

Example of Networks - The WWW

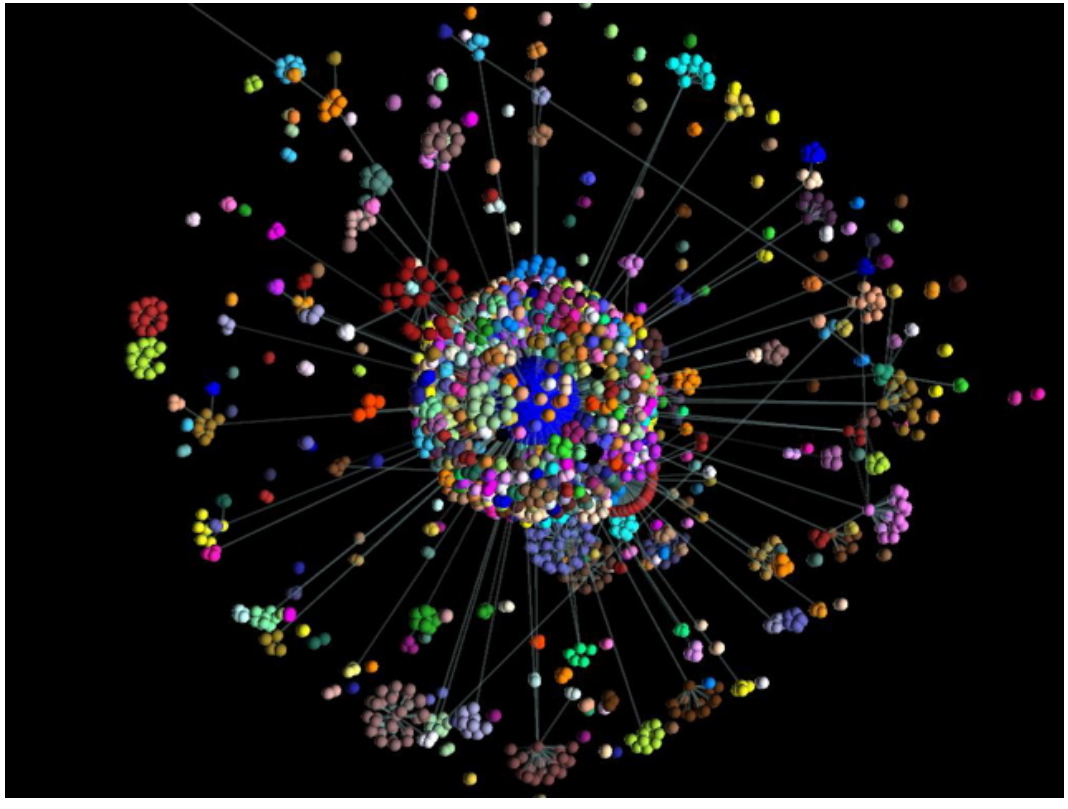


Figure: source by: www.http://thesituationist.wordpress.com/2008/03/03/social-networks/

Example of Networks - Obesity

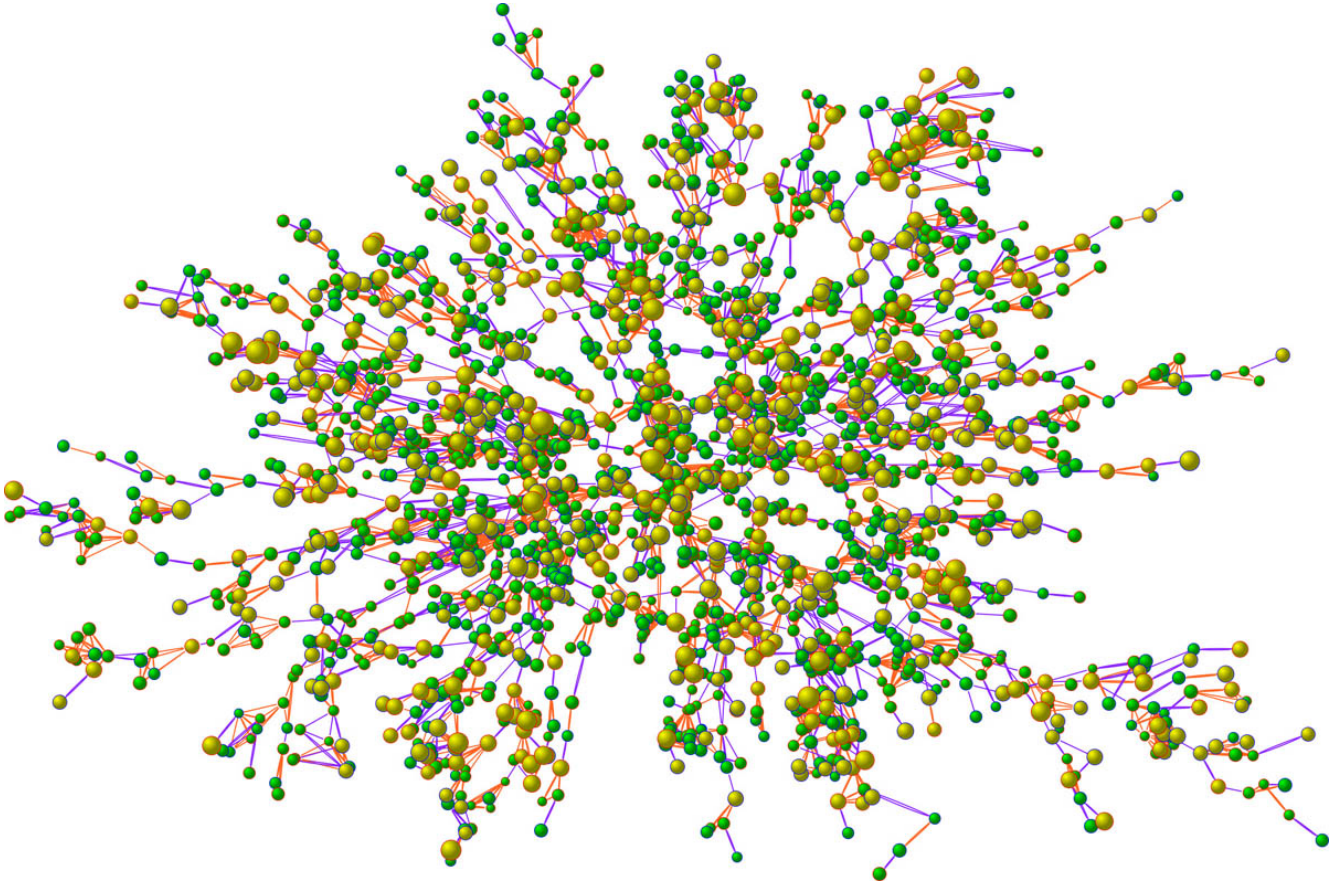
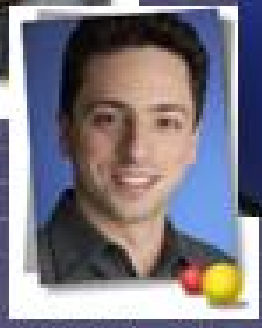
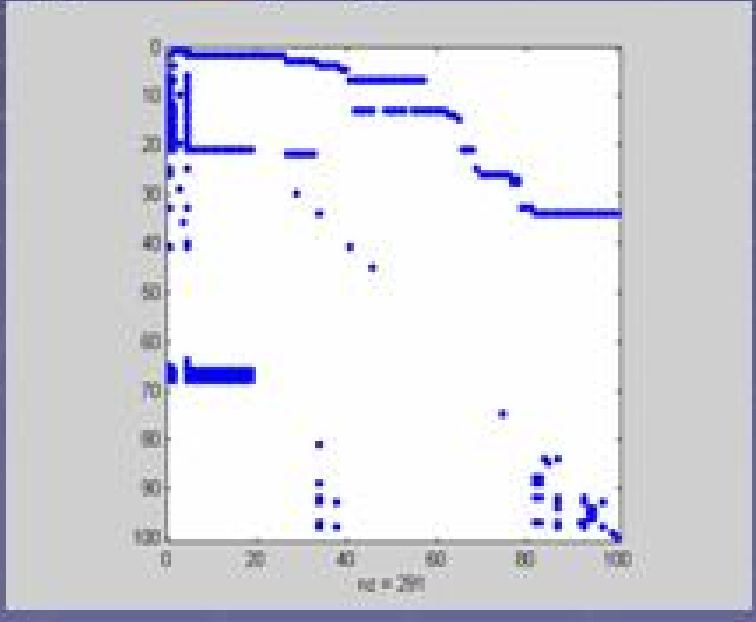
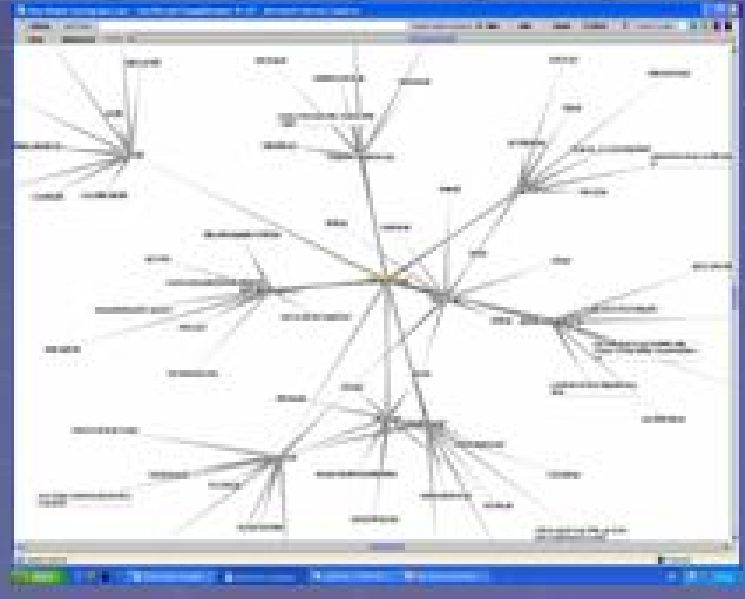


Figure: source by: http://lev.ccny.cuny.edu/~hmakse/complexnets.html



L. Page & S. Brin (Google)





Η Google και η βαθμολόγηση PageRank

Computer Networks and ISDN Systems 30 (1998) 107–117

The anatomy of a large-scale hypertextual Web search engine [1]

Sergey Brin [2], Lawrence Page [*,2]

Computer Science Department, Stanford University, Stanford, CA 94305, USA

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/

To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of Web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the Web, very little academic research has been done on them. Furthermore, due to rapid advance in technology and Web proliferation, creating a Web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale Web search engine — the first such detailed public description we know of to date.

Apart from the problems of scaling traditional search techniques to data of this magnitude, there are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with uncontrolled hypertext collections where anyone can publish anything they want.

Keywords: World Wide Web; Search engines; Information retrieval; PageRank; Google

1. Introduction

The Web creates new challenges for information retrieval. The amount of information on the Web is growing rapidly, as well as the number of new users inexperienced in the art of Web research. People are likely to surf the Web using its link graph, often starting with high quality human maintained indices such as **Yahoo!** [3] or with search engines. Human maintained lists cover popular topics effectively but are subjective, expensive to build and maintain, slow to improve, and cannot cover all esoteric topics. Automated search engines that rely on keyword matching usually return too many low quality matches. To make matters worse, some advertisers attempt to gain people's attention by taking measures meant to mislead

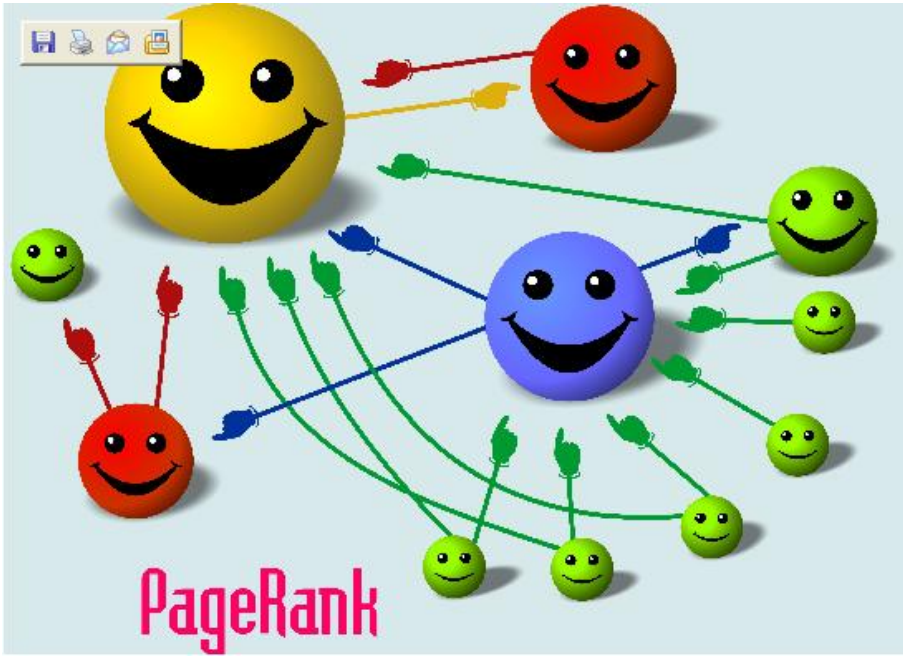
* Corresponding author.
[1] There are two versions of this paper — a longer full version and a shorter printed version. The full version is available on the Web and the conference CD-ROM.
[2] E-mail: {sergey, page}@cs.stanford.edu
[3] http://www.yahoo.com/

0169-7552/98/$19.00
PII S0169-7552(98)00110-X

Στόχοι της Google

Κατά τον Larry Page η τέλεια μηχανή αναζήτησης είναι εκείνη που `καταλαβαίνει ακριβώς τι εννοείς και επιστρέφει ακριβώς αυτό που χρειάζεσαι΄.

Βασικοί συντελεστές: Συνάφεια - Φρεσκότητα - Πληρότητα - Ταχύτητα



Συνάφεια - Πληρότητα - Ταχύτητα

... μια σημαντική πρωτοτυπία ήταν το PageRank, ένας αλγόριθμος που βαθμολογεί/κατατάσσει κάθε ιστοσελίδα εξετάζοντας ποιές ιστοσελίδες δείχνουν σ' αυτήν, καθώς και άλλα δεδομένα. Η δεικτοδοτεί περισσότερες από $O(10^9)$ ιστοσελίδες ενώ ο μέσος χρόνος για μια απάντηση είναι περί τα 0.25sec.

THE $25,000,000,000* EIGENVECTOR
THE LINEAR ALGEBRA BEHIND GOOGLE

KURT BRYAN† AND TANYA LEISE‡

Abstract. Google's success derives in large part from its PageRank algorithm, which ranks the importance of webpages according to an eigenvector of a weighted link matrix. Analysis of the PageRank formula provides a wonderful applied topic for a linear algebra course. Instructors may assign this article as a project to more advanced students, or spend one or two lectures presenting the material with assigned homework from the exercises. This material also complements the discussion of Markov chains in matrix algebra. Maple and Mathematica files supporting this material can be found at www.rose-hulman.edu/~bryan.

Key words. linear algebra, PageRank, eigenvector, stochastic matrix

AMS subject classifications. 15-01, 15A18, 15A51

Ίσως ο μεγαλύτερος υπολογισμός με μητρώα στον κόσμο!

12 CLEVE'S CORNER

THE WORLD'S LARGEST MATRIX COMPUTATION

Google's PageRank is an eigenvector of a matrix of order 2.7 billion.

BY CLEVE MOLER

One of the reasons why Google is such an effective search engine is the PageRank™ algorithm, developed by Google's founders, Larry Page and Sergey Brin, when they were graduate students at Stanford University. PageRank is determined entirely by the link structure of the Web. It is recomputed about once a month and does not involve any of the actual content of Web pages or of any individual query. Then, for any particular query, Google finds the pages on the Web that match that query and lists those pages in the order of their PageRank.

Imagine surfing the Web, going from page to page by randomly choosing an outgoing link from one page to get to the next. This can lead to dead ends at pages with no outgoing links, or cycles around cliques of interconnected pages. So, a certain fraction of the time, simply choose a random page from anywhere on the Web. This theoretical random walk of the Web is a *Markov chain* or *Markov process*. The limiting probability that a dedicated random surfer visits any particular page is its PageRank. A page has high rank if it has links to and from other pages with high rank.

Let W be the set of Web pages that can reached by following a chain of hyperlinks starting from a page at Google and let n be the number of pages in W. The set W actually varies with time, but in May 2002

It tells us that the largest eigenvalue of A is equal to one and that the corresponding eigenvector, which satisfies the equation

$$x = Ax,$$

exists and is unique to within a scaling factor. When this scaling factor is chosen so that

$$\sum_i x_i = 1$$

then x is the state vector of the Markov chain. The elements of x are Google's PageRank.

If the matrix were small enough to fit in MATLAB, one way to compute the eigenvector x would be to start with a good approximate solution, such as the PageRanks from the previous month, and simply repeat the assignment statement

```
x = Ax
```

until successive vectors agree to within specified tolerance. This is known as the power method and is about the only possible approach for very large n. I'm not sure how Google actually computes PageRank, but one step of the power method would require one pass over a database of Web pages, updating weighted reference

Ίσως ο μεγαλύτερος υπολογισμός με μητρώα στον κόσμο!

SIAM REVIEW
Vol. 48, No. 3, pp. 569–581

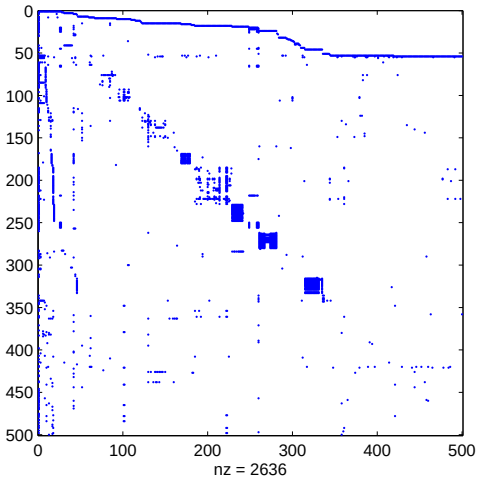
The $25,000,000,000 Eigenvector: The Linear Algebra behind Google*

Kurt Bryan†
Tanya Leise‡

Abstract. Google's success derives in large part from its PageRank algorithm, which ranks the importance of web pages according to an eigenvector of a weighted link matrix. Analysis of the PageRank formula provides a wonderful applied topic for a linear algebra course. Instructors may assign this article as a project to more advanced students or spend one or two lectures presenting the material with assigned homework from the exercises. This material also complements the discussion of Markov chains in matrix algebra. Maple and *Mathematica* files supporting this material can be found at www.rose-hulman.edu/∼bryan.

Key words. linear algebra, PageRank, eigenvector, stochastic matrix

Γράφημα $\rightarrow$ Μητρώο γειτνίασης



Μητρώο γειτνίασης A

$$A_{ij} = \begin{cases} 1, & i \text{ δείχνει τη σελίδα } j \\ 0, & \text{διαφορετικά} \end{cases}$$

Harvard500

1	http://www.harvard.edu
2	http://atwork.harvard.edu
3	http://lib.harvard.edu
...	...
500	http://www.hsdm.med.harvard.edu/(...)/implant.htm

Μητρώο μετάβασης & στοχαστικοποίηση (Brin & Page)

Μητρώο μετάβασης P

$$P_{ij} = \begin{cases} \dfrac{A_{ij}}{\deg(i)}, & \text{αν } \deg(i) \neq 0 \text{ όπου } \deg(i) = \sum_j A_{ij} \quad (1) \\ 0, & \text{διαφορετικά} \quad (2) \end{cases}$$

1η Επιδιόρθωση: Αποφυγή προβλημάτων στις καταβόθρες.

$$S = P^\top + w\, d^T, \quad w = \frac{1}{n} e$$

2η Επιδιόρθωση: Για μοναδικότητα πρέπει να υπάρχει μοναδική μέγιστη ιδιοτιμή $= 1$. Επιλέγουμε στοχαστικό διάνυσμα $v \geq 0$, οπότε αν $\mu \in (0,1)$ αποδεικνύεται ότι υπάρχει μοναδικό στοχαστικό διάνυσμα $p \geq 0$ τέτοιο ώστε $G(\mu)p = p$. Αν επιπλέον το $G(\mu)$ είναι μη αναγωγήσιμο, ισχύει ότι $p > 0$.

Παραμετροποιημένο μητρώο Google

$$G(\mu) = \mu S + (1-\mu) v e^\top$$

Διάταξη PageRank

Ερμηνεία μέσω του εργοδικού θεωρήματος

Το $G(\mu)$ είναι στοχαστικό κατά στήλες και υπάρχει μόνο μία πραγματική ιδιοτιμή με μέτρο 1, επομένως

$$\lim_{k\to\infty} G(\mu)^k = pe^\top$$

Το διάνυσμα p είναι στοχαστικό[1] και λέγεται διάνυσμα PageRank. Η κατάταξη PageRank προκύπτει άμεσα από τη διάταξη των τιμών του $p > 0$.

Παρατηρήσεις

- Δεν είναι πρακτική μέθοδος για να υπολογίσουμε το PageRank
- Αν $0 < \mu < 1$ αποδεικνύεται ότι η ιδιοτιμή 1 είναι μοναδική και σε απόλυτη τιμή, και ότι $p > 0$.

[1] Είναι δηλαδή το διάνυσμα Perron του $G(\mu)$.

Πρακτικοί χαρακτηρισμοί του διανύσματος κατάταξης ιστοσελίδων

Ως ιδιοδιάνυσμα:

το διάνυσμα PageRank x^{PR} είναι το στοχαστικό διάνυσμα που ικανοποιεί τη σχέση

$$x = G\,x$$

- το x είναι το θετικό ιδιοδιάνυσμα που αντιστοιχεί στο $\lambda_1 := \max \lambda(G) = 1$.

Ως λύση γραμμικού συστήματος:

το διάνυσμα PageRank x^{PR} ικανοποιεί τη σχέση

$$(I - \mu S)\,x = (1-\mu)\,v, \quad x > 0, \|x\|_1 = 1.$$

Εναλλακτικός χαρακτηρισμός

Με τυχαίους περιπατητές[2]

Πρόκειται για ιδεατούς χρήστης που κάνουν τυχαίους περιπάτους στις ιστοσελίδες. Κάθε πλοηγητής, από κάθε σελίδα, με πιθανότητα μ επιλέγει ισοπίθανα κάποιον από τους εξερχόμενους συνδέσμους και με πιθανότητα $1 - \mu$ επιλέγει ισοπίθανα οποιονδήποτε άλλον. Αν πολλοί τυχαίοι πλοηγητές αρχίσουν να «περπατούν» στο γράφημα (του Παγκόσμιου Ιστού) τότε μετά από ένα χρονικό διάστημα, το πλήθος των πλοηγητών που βρίσκεται σε κάθε κόμβο καθορίζει και το βαθμό PageRank κάθε ιστοσελίδας.

Προσοχή: Ο PageRank είναι ένα στοιχείο, από τα πολλά, που χρησιμοποιεί η Google για να μετρήσει τη σημαντικότητα των ιστοσελίδων.

ΠΡΟΣΟΧΗ: Μόνον η Google γνωρίζει την ακριβή «συνταγή»!

[2] Πιο πετυχημένος ο αμετάφραστος όρος random surfers

Int J Parallel Prog
DOI 10.1007/s10766-014-0344-3

PageRank Computation Using a Multiple Implicitly Restarted Arnoldi Method for Modeling Epidemic Spread

Zifan Liu · Nahid Emad · Soufian Ben Amor · Michel Lamure

Received: 5 January 2014 / Accepted: 24 October 2014

Distributing antidote using PageRank vectors

Fan Chung*
University of California, San Diego

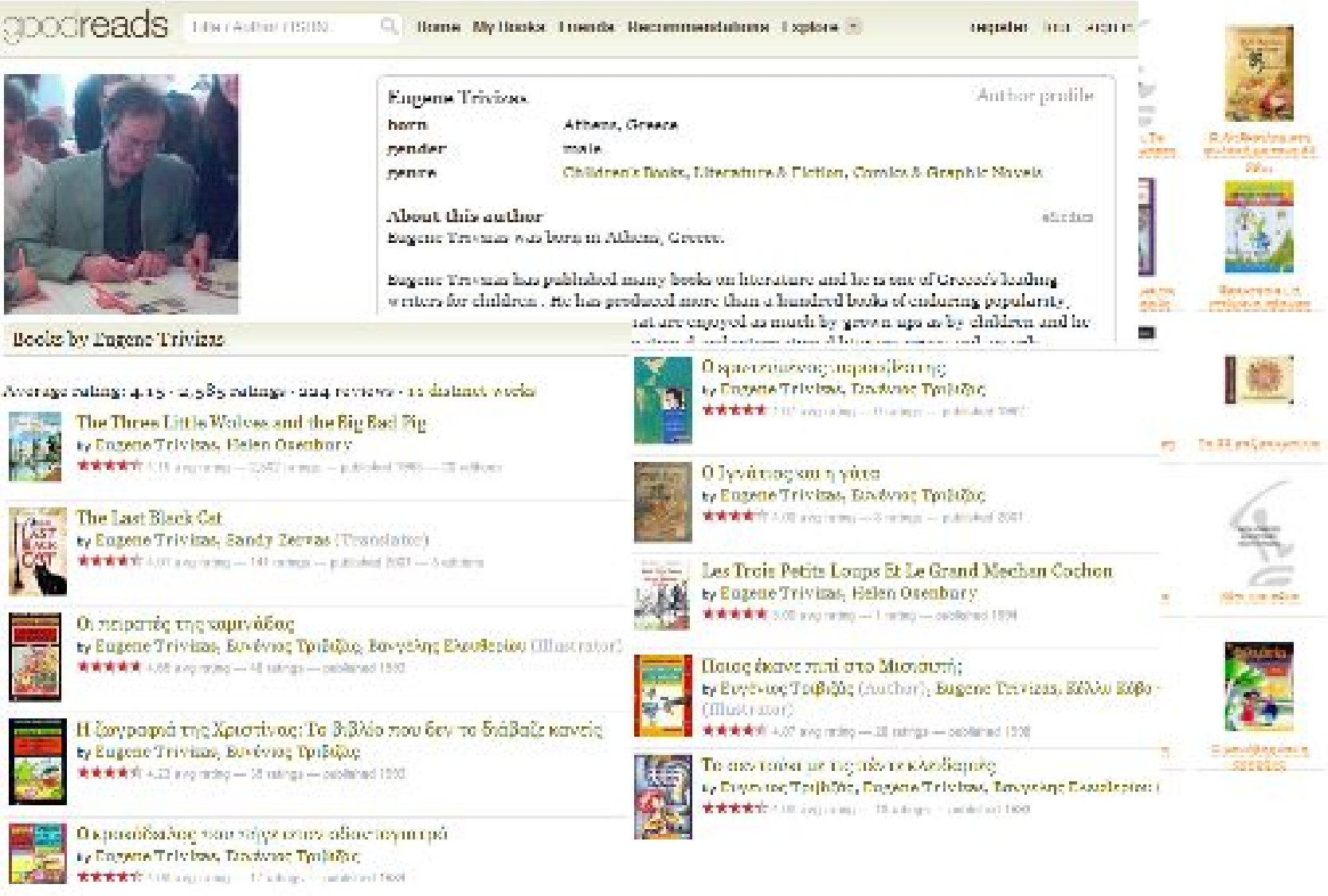
Paul Horn
Emory University

Alexander Tsiatas
University of California, San Diego

Abstract

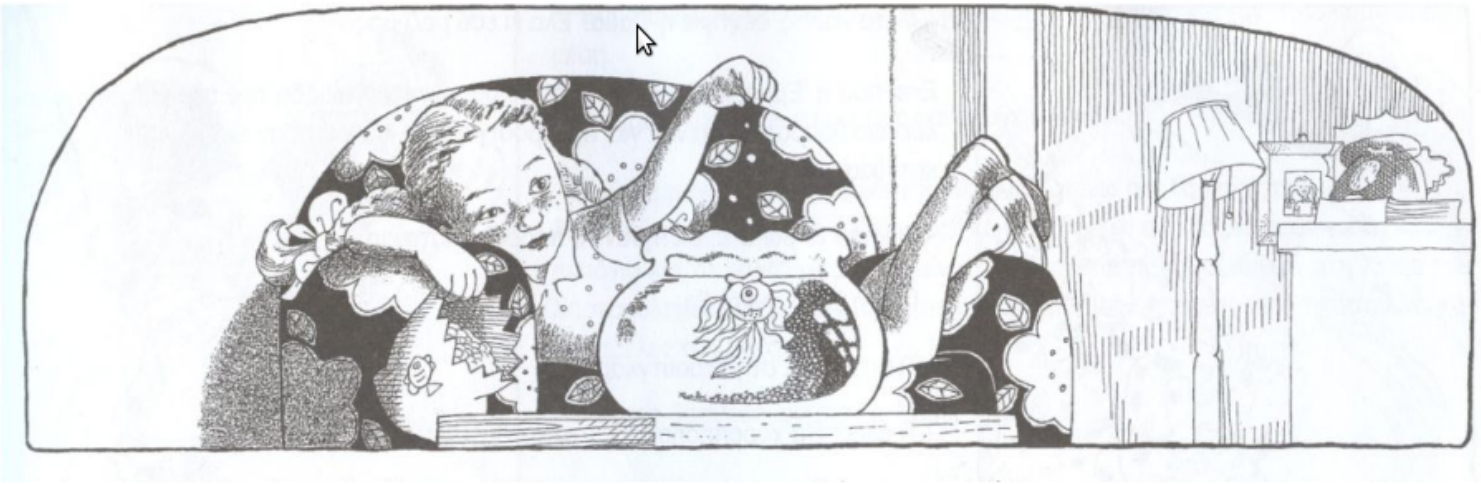
We give an analysis of a variant of the contact process on finite graphs, allowing for non-uniform cure rates, modeling antidote distribution. We examine an inoculation scheme using PageRank vectors which quantify the correlations among vertices in the contact graph. We show that for a contact graph on n nodes we can select a set H of nodes to inoculate such that with probability at least $1-2\epsilon$, any infection from any starting infected set of s nodes will die out in $c \log s + c'$ time, where c and c' depend only on the probabilitstic error bound ϵ and the infection rate, and the size of H depends only on s, ϵ and the topology around the initially infected nodes, independent of the size of the whole graph.

The world of Eugenios Trivizas



The *88 dolmadakia*

Start a reading (from p7)



Once upon a time in a country house up on 33 Onestroke road (never foget this, it's important) lived a girl with freckles and red pigtails whose name was Emma. Emma had one mother, twelve aunts and one beautiful pink goldfish. One morning that Emma was alone in the house and was feeding the goldfish suddenly …. ring ring … the phone rings. Emma stood a little undecisive.

7

WHAT WOULD YOU LIKE EMMA TO DO?

ANSWER! → *Read the sequel on page 55*

NOT ANSWER! → *Read the sequel on page 101*

Emma did not answer so the phone morphed into a pot full of chrysanthema. Suddenly four firemen with red mustaches and blue watering cans entered from the open windows and watered the plant. The chrysanthema started growing, so quickly that they soon pierced the ceiling and were no longer seen as they had already reached the clouds, the sky and beyond.
Emma climbed onto the tallest branch and when she reached cloud nine she started hitch-hiking. Soon enough a multicolor balloon with a gondola stopped near her. Inside the gondola a plump cook with rosy cheeks and wearing a white chef cap with blue patches was jumping around.

"Come aboard!" the cook told Emma. Fast! We have no time to lose!
Emma hopped in the gondola! "Where are we going?" she asked.

WHERE WOULD YOU LIKE THEM TO GO? **TO** THE BIRTHDAY OF THE KING OF DOLMANDIA **OR** TO THE RECEPTION OF THE KING OF YAMMISTAN?

7

101

DOLMANDIA'S KING BIRTHDAY! → *Continue on the page whose first digit is found w h e r e t o ends and the second digit is hidden in a x i s.*

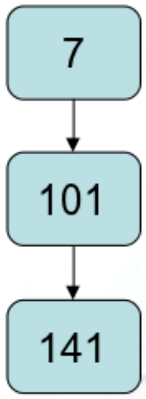
YAMMISTAN'S KING RECEPTION! → *Continue on the page whose first two digits are twice as many as the dwarfs in the tale of Whitesnow and the third is as many as the eyes of a Cyclop.*

To the reception of Yammistan's king. He invited me to cook potato dumplings for his guests. Shortly afterwards the balloon landed in the king's palace garden. The garden was paved with bricks, all pink and blue.

– You are late, the king exclaimed! I have invited ten harlequins and I want you to cook three potato dumplings for each. Beware because if you cook too few, I will have the sorcerer turn you into geese! And if you cook too many, he will turn you into slippers!

Emma and the cook ran to the kitchen.

– Please tell me, the cook asked Emma. How many potato dumplings must we cook so that the king is happy and the sorcerer leaves us alone?

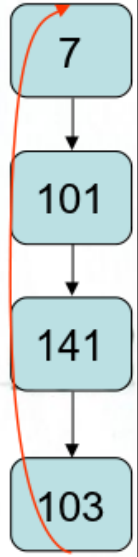


WHAT SHOULD EMMA RESPOND?

TEN POTATO DUMPLINGS! *Read the sequel on page 118*

THIRTY DUMPLINGS! *Read the sequel on page 103*

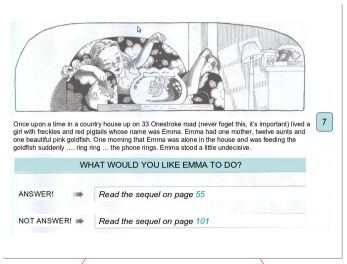
SEVENTY THOUSAND! *Read the sequel on page 114*



They prepared thirty potato dumplings to have exactly three for the ten harlequins to eat. The king was so happy that he requested that the cook marry his daughter, princess Sophia, and offered Emma one candy apple, one bird, one silver walnut and a flying basket. Emma got into the basket and flew home where she lived happily thereafter.

THE END

If you wish, you can read the story again from the beginning, this time making different choices.



Once upon a time in a country house up on 33 Onestroke road (never forget this, it's important) lived a girl with freckles and red pigtails whose name was Emma. Emma had one mother, twelve aunts and one beautiful pink goldfish. One morning that Emma was alone in the house and was feeding the goldfish suddenly .... ring ring ... the phone rings. Emma stood a little undecisive.

7

WHAT WOULD YOU LIKE EMMA TO DO?

ANSWER! → *Read the sequel on page 55*

NOT ANSWER! → *Read the sequel on page 101*

page55

Emma did not answer so the phone morphed into a pot full of chrysanthema. Suddenly four firemen with red mustaches and blue watering cans entered from the open windows and watered the plant. The chrysanthema started growing, so quickly that they soon pierced the ceiling and were no longer seen as they had already reached the clouds, the sky and beyond. Emma climbed onto the tallest branch and when she reached cloud nine she started hitch-hiking. Soon enough a multicolor balloon with a gondola stopped near her. Inside the gondola a plump cook with rosy cheeks and wearing a white chef cap with blue patches was jumping around.

"Come aboard!" the cook told Emma. Fast! We have no time to lose!
Emma hopped in the gondola! "Where are we going?" she asked.

7 → 101

WHERE WOULD YOU LIKE THEM TO GO? **TO** THE BIRTHDAY OF THE KING OF DOLMANDIA **OR** TO THE RECEPTION OF THE KING OF YAMMISTAN?

DOLMANDIA'S KING BIRTHDAY! → *Continue on the page whose first digit is found w h e r e t o ends and the second digit is hidden in a x i s.*

YAMMISTAN'S KING RECEPTION! → *Continue on the page whose first two digits are twice as many as the dwarfs in the tale of Whitesnow and the third is as many as the eyes of a Cyclop.*

page67 page38 page133 page97

To the reception of Yammistan's king. He invited me to cook potato dumplings for his guests. Shortly afterwards the balloon landed in the king's palace garden. The garden was paved with bricks, all pink and blue.
– You are late, the king exclaimed! I have invited ten harlequins and I want you to cook three potato dumplings for each. Beware because if you cook too few, I will have the sorcerer turn you into geese! And if you cook too many, he will turn you into slippers!
Emma and the cook ran to the kitchen.
– Please tell me, the cook asked Emma. How many potato dumplings must we cook so that the king is happy and the sorcerer leaves us alone?

7 → 101 → 141

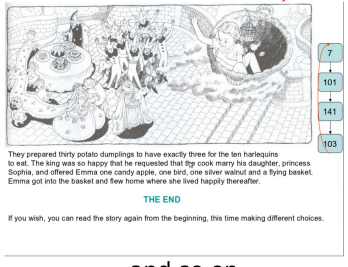
WHAT SHOULD EMMA RESPOND?

TEN POTATO DUMPLINGS! → *Read the sequel on page 118*

THIRTY DUMPLINGS! → *Read the sequel on page 103*

SEVENTY THOUSAND! → *Read the sequel on page 114*

page26



They prepared thirty potato dumplings to have exactly three for the ten harlequins to eat. The king was so happy that he requested that the cook marry his daughter, princess Sophia, and offered Emma one candy apple, one bird, one silver walnut and a flying basket. Emma got into the basket and flew home where she lived happily thereafter.

THE END

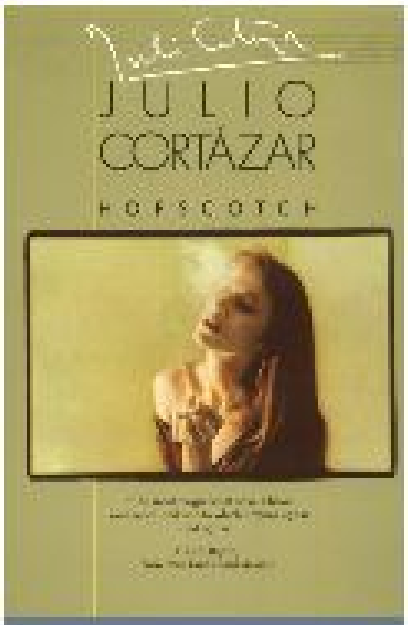
If you wish, you can read the story again from the beginning, this time making different choices.

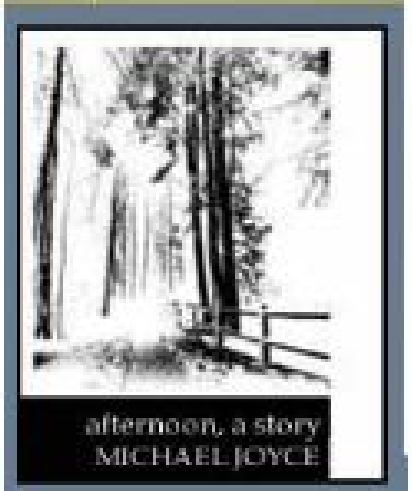
7 → 101 → 141 → 103

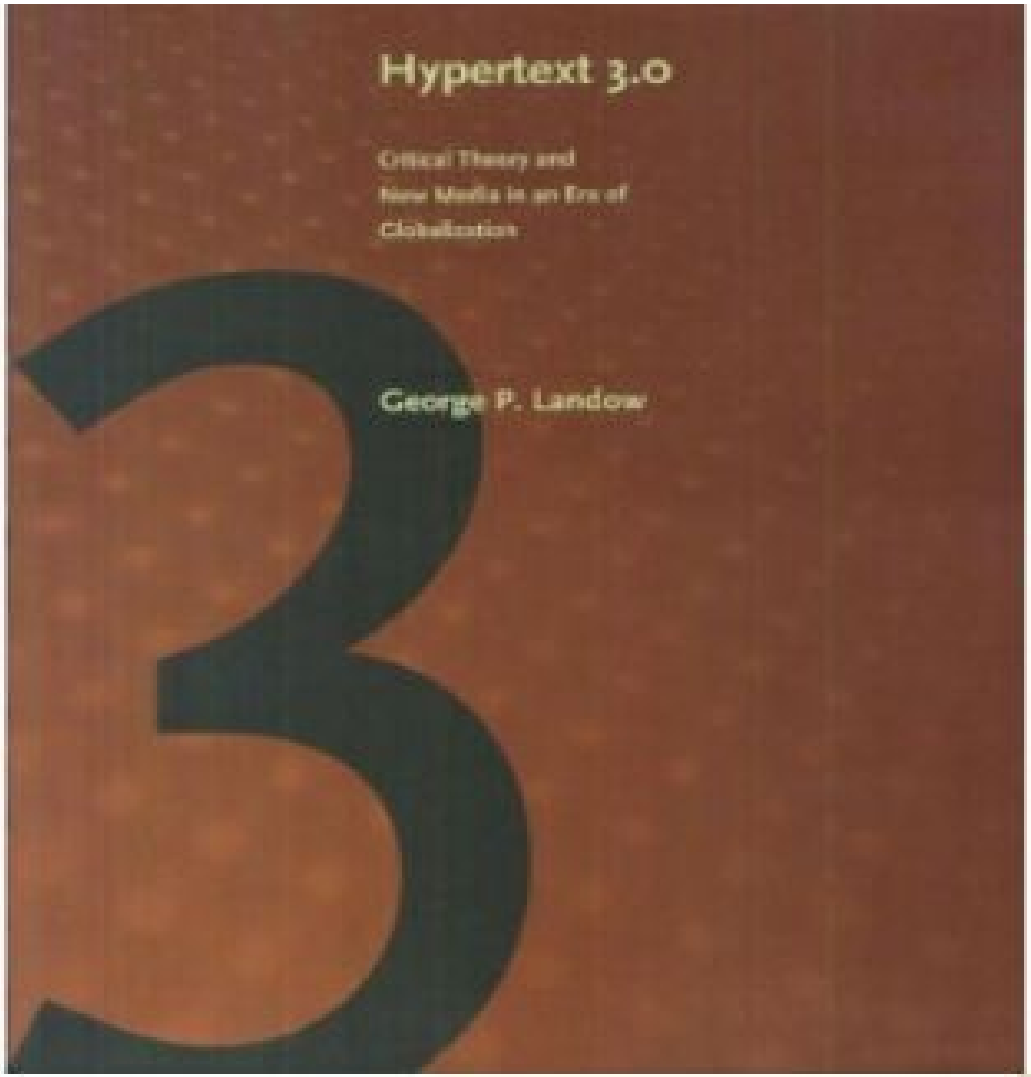
page118 page114

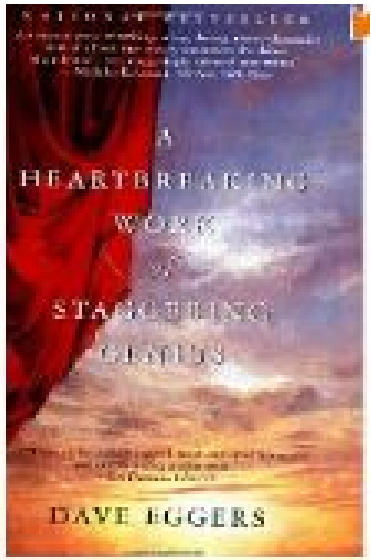
and so on...

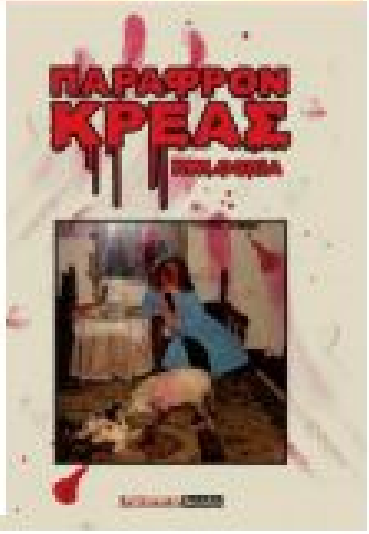
Interest


JULIO
CORTÁZAR
RAYUELA

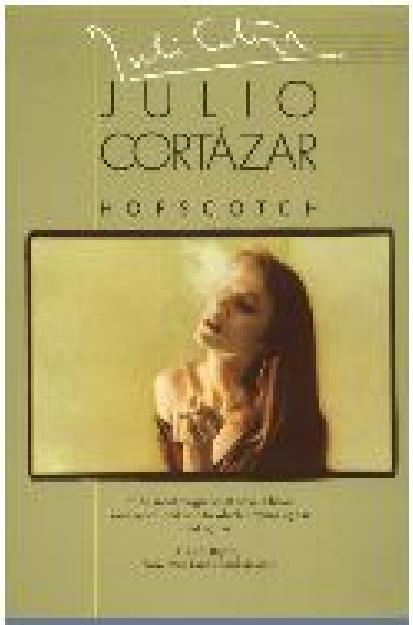

afternoon, a story
MICHAEL JOYCE

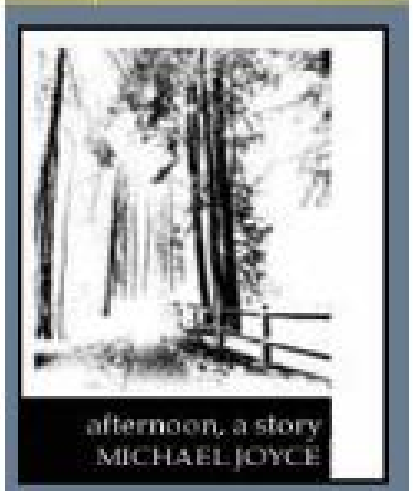

Hypertext 3.0
Critical Theory and
New Media in an Era of
Globalization
George P. Landow


A
HEARTBREAKING
WORK
OF
STAGGERING
GENIUS
DAVE EGGERS


ΠΑΡΑΦΡΟΝ
ΚΡΕΑΣ

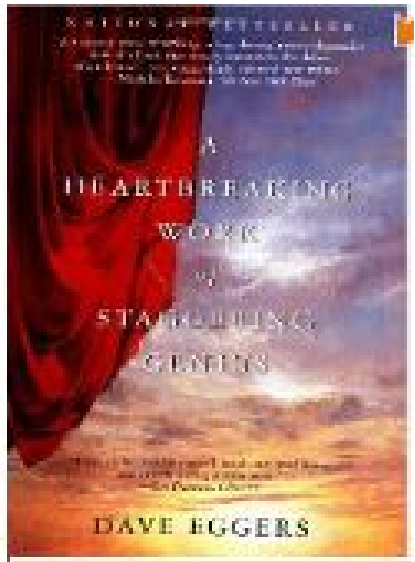
Interest?

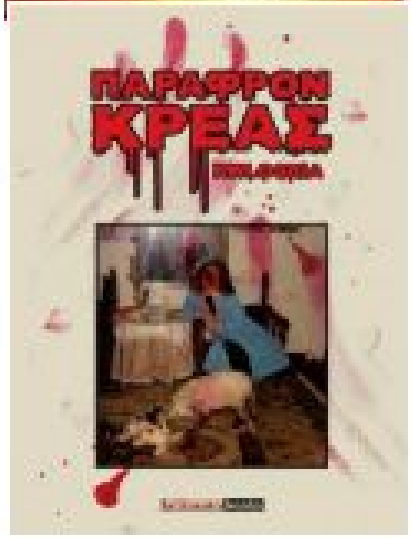




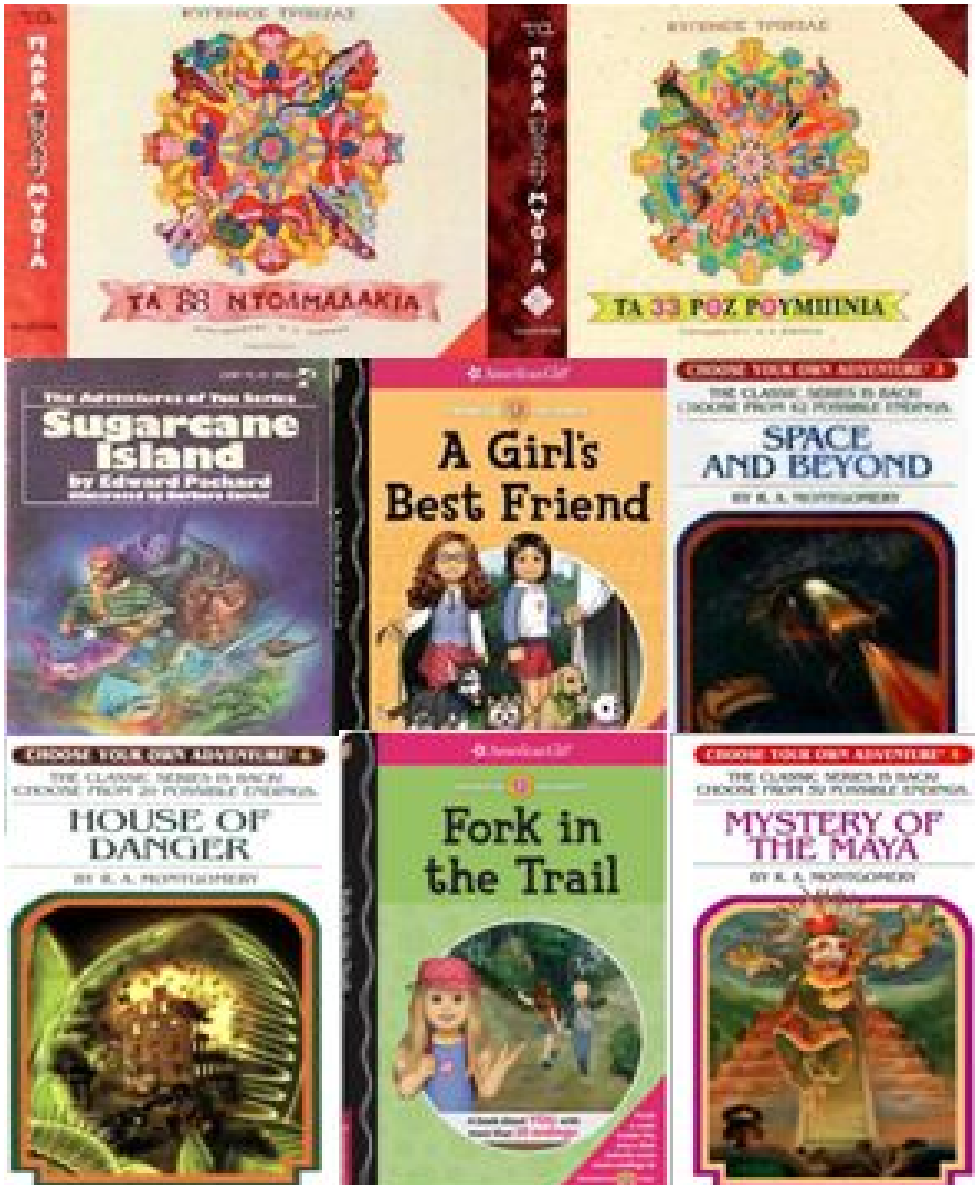
Indeed, a rich realm of possibilities appears to be open. We would like to explore them, bearing in mind, however, these words of caution: Of course, what the reader or rather the user gains in possibilities, he or she probably loses in aesthetic pleasure.

Is this freedom something to be hoped for? **Freedom is a great thing in life. But enjoyment of art, narrative art more especially, is a field of human activity where we delight in having less, not more, choices. It is no wonder then that hypertext fiction did not really catch on....** [Doxiadis A., (2005)]





Interesting for kids



The French connection

Queneau, Raymond [1967].
***Un conte à votre façon*, in Oulipo. *La littérature potentielle (Créations Re-créations Récréations)*. Paris: Gallimard, coll. «Folio|Essais», 1973, pp.273-276.**

1. Désirez-vous connaître l'histoire des trois alertes petits pois?
   a. si oui, passez à 4
   b. si non, passez à 2
2. Préférez-vous celle des trois minces grands échalas?
   a. si oui, passez à 16
   b. si non, passez à 3
3. Préférez-vous celle des trois moyens médiocres arbustes?
   a. si oui, passez à 17
   b. si non, passez à 21
4. Il y avait une fois trois petits pois vêtus de vert qui dormaient gentiment dans leur cosse. Leur visage bien rond respirait par les trous de leurs narines et l'on entendait leur ronflement doux et harmonieux.
   a. si vous préférez une autre description, passez à 9
   b. si celle-ci vous convient, passez à 5
5. Ils ne rêvaient pas. Ces petits êtres en effet ne rêvent jamais.
   a. si vous préférez qu'ils rêvent, passez à 6
   b. sinon, passez à 7
6. Ils rêvaient. Ces petits êtres en effet rêvent toujours et leurs nuits sécrètent des songes charmants.
   a. si vous désirez connaître ces songes passez à 11
   b. si vous n'y tenez pas, vous passez à 7
7. Leurs pieds mignons trempaient dans de chaudes chaussettes et ils portaient au lit des gants de velours noir.
   a. si vous préférez des gants d'une autre couleur passez à 8
   b. si cette couleur vous convient, passez à 10.
8. Ils portaient au lit des gants de velours bleu.
   a. si vous préférez des gants d'une autre couleur, passez à 7
   b. si cette couleur vous convient, passez à 10
9. Il y avait une fois trois petits pois qui roulaient leur bosse sur les grands chemins. Le soir venu, fatigués et las, ils s'endormirent très rapidement.
   a. si vous désirez connaître la suite, passez à 5
   b. si non, passez à 21
10. Tous les trois faisaient le même rêve, ils s'aimaient en effet tendrement et, en bons fiers trumeaux, songeaient toujours semblablement.
    a. si vous désirez connaître leur rêve, passez à 11
    b. si non, passez à 12
11. Ils rêvaient qu'ils allaient chercher leur soupe à la cantine populaire et qu'en ouvrant leur gamelle ils découvraient que c'était de la soupe d'ers. D'horreur, ils se réveillaient.
    a. si vous voulez savoir pourquoi ils s'éveillent d'horreur, consultez le Larousse au mot «ers» et n'en parlons plus
    b. si vous jugez inutile d'approfondir la question, passez à 12
12. Opopoï! s'écrient-ils en ouvrant les yeux. Opopoï! Quel songe avons-nous enfanté là! Mauvais présage, dit le premier. Oui-da, dit le second, c'est bien vrai, me voilà triste. Ne vous troublez pas ainsi, dit le troisième qui était le plus futé, il ne s'agit pas de s'émouvoir, mais de comprendre, bref je m'en vais vous analyser ça.
    a. si vous désirez connaître tout de suite l'interprétation de ce songe, passez à 15
    b. si vous souhaitez au contraire connaître les réactions des deux autres, passez à 13
13. Tu nous la bailles belle, dit le premier. Depuis quand sais-tu analyser les songes? Oui, depuis quand sais-tu analyser les songes? Oui, depuis quand? ajouta le second.
    a. si vous désirez aussi savoir depuis quand, passez à 14
    b. si non, passez à 14 tout de même, car vous ne le saurez pas plus
14. Depuis quand? s'écria le troisième. Est-ce que je sais moi! Le fait est que je pratique la chose. Vous allez voir!
    a. si vous voulez vous aussi voir, passez à 15
    b. si non, passez également à 15, car vous ne verrez rien.
15. Eh bien! voyons, dirent ses frères. Votre ironie ne me plaît pas, répliqua l'autre, et vous ne saurez rien. D'ailleurs, au cours de cette conversation d'un ton assez vif, votre sentiment d'horreur ne s'est-il pas estompé? effacé même? Alors quoi bon remuer le bourbier de votre inconscient de papilionacées? Allons plutôt nous laver à la fontaine et saluer ce gai matin dans l'hygiène et la sainte euphorie! Aussitôt dit, aussitôt fait: les voilà qui se glissent hors de leur cosse, se laissent doucement rouler sur le sol et puis au petit trot gagnent joyeusement le théâtre de leurs ablutions.
    a. si vous désirez savoir ce qui se passe sur le théâtre de leurs ablutions, passez à 16
    b. si vous ne le désirez pas, vous passez à 21
16. Trois grands échalas les regardaient faire.
    a. si les trois échalas vous déplaisent, passez à 21
    b. s'ils vous conviennent, passez à 18
17. Trois moyens médiocres arbustes les regardaient faire.
    a. si les trois moyens médiocres arbustes vous déplaisent, passez à 21
    b. s'ils vous conviennent, passez à 18
18. Se voyant ainsi zyeutés, les trois alertes petits pois qui étaient fort pudiques s'ensauvèrent.
    a. si vous désirez savoir ce qu'ils firent ensuite, passez à 19
    b. si vous ne le désirez pas, vous passez à 21
19. Ils coururent bien fort pour regagner leur cosse et, refermant celle-ci derrière eux, s'y endormirent de nouveau.
    a. si vous désirez connaître la suite, passez à 20
    b. si vous ne le désirez pas, vous passez à 21
20. Il n'y a pas de suite le conte est terminé.
21. Dans ce cas, le conte est également terminé.

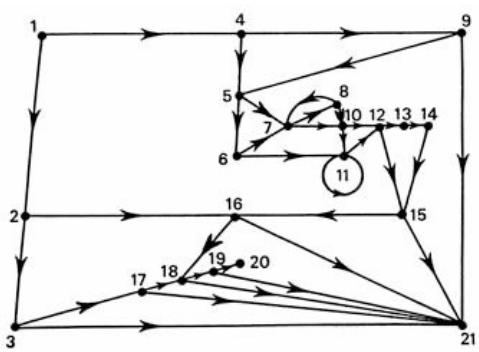
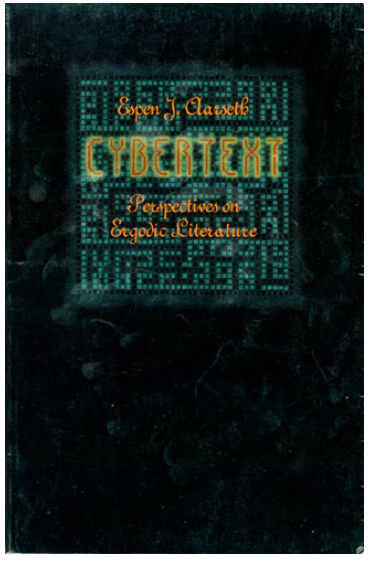


Figure 1 – Graphe du «Conte à votre façon»

Ergodic Literature (Aarseth)



Definition (Aarseth´97)

During the cybertextual process, the user will have effectuated a semiotic sequence, and this selective movement is a work of physical construction that the various concepts of "reading" do not account for. This phenomenon I call ergodic, using a term appropriated from physics that derives from the Greek words ergon and hodos, meaning "work" and "path."
In ergodic literature, nontrivial effort is required to allow the reader to traverse the text.

Structure

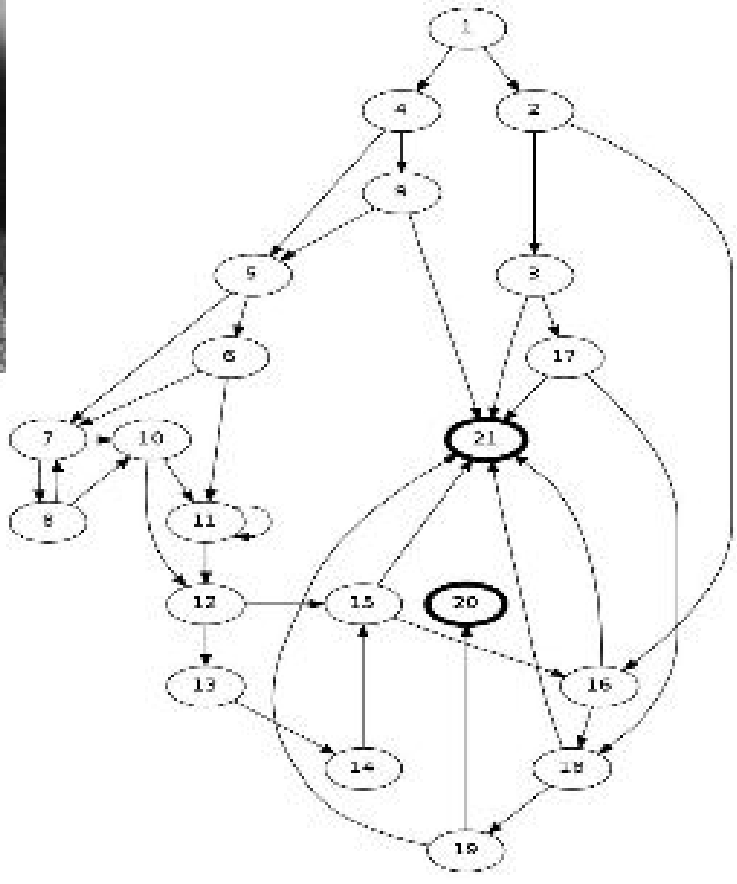
Storylet

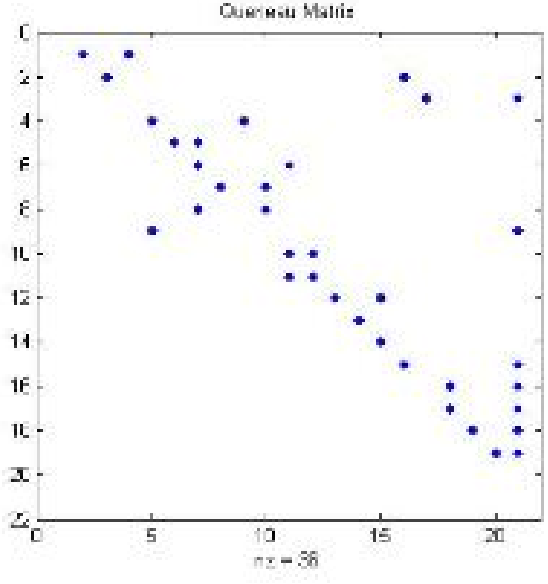
Textual or visual material contained within a single page terminated by ``branch'' or ``end''. Storylets can be ``starting'', ``ending'' or ``intermediate''.

$$\text{Plot} \approx \text{Links} + \text{Storylets}$$

- special case of ``lexias'' (Barthes'70), substories, etc.
- fusion flavor ``children's tales meet computer science''

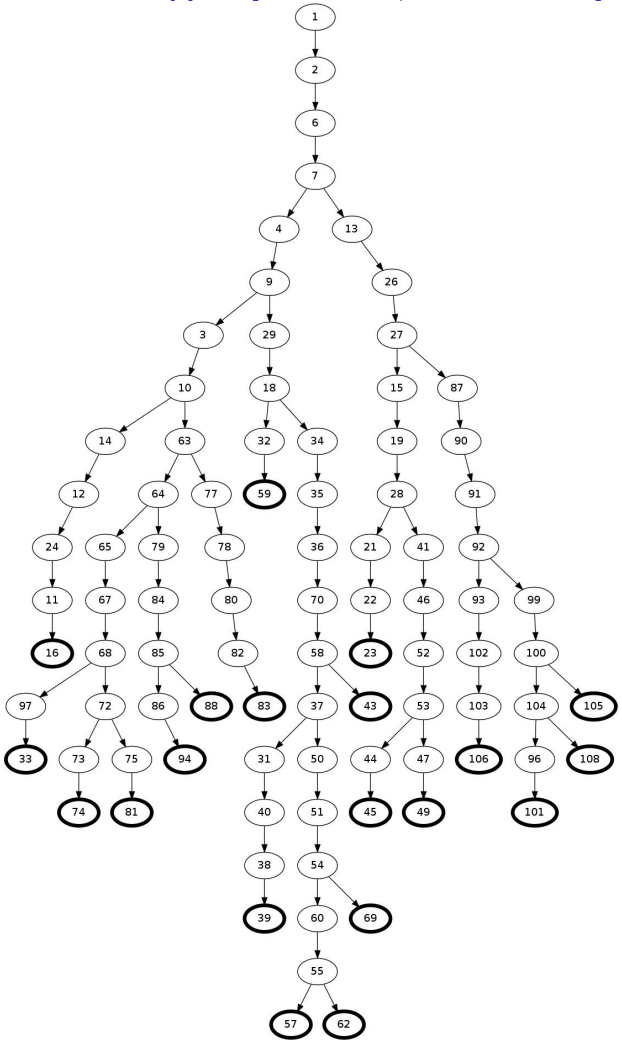
From IF books to digraphs and matrices

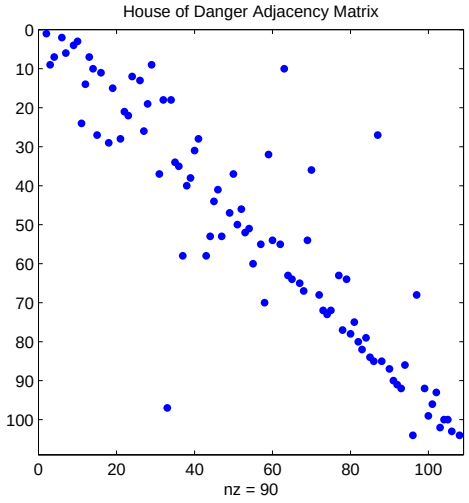




Range from quite simple ...

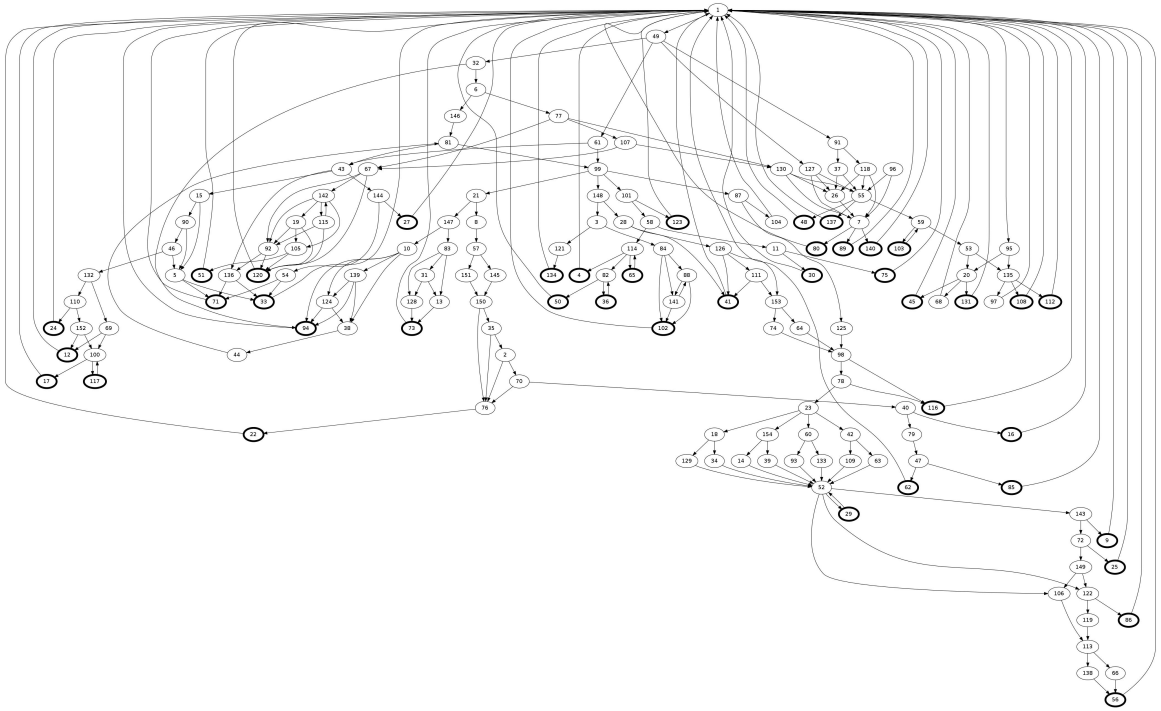
House of Danger ($n = 91, \mathtt{nnz} = 90$)





... to more complicated

33 Pink Rubies ($n = 211, \texttt{nnz} = 386$)



``My daughter and I began this book when she was three and got bored when she was seven. Incredible imagination[3] *''*

[3] www.greekbooks.gr/books/pedika/pedika/ta-33-roz-rubinia.product

Maths and Graphs

What can graph and matrix based mathematical models tell us about ergodic literature?

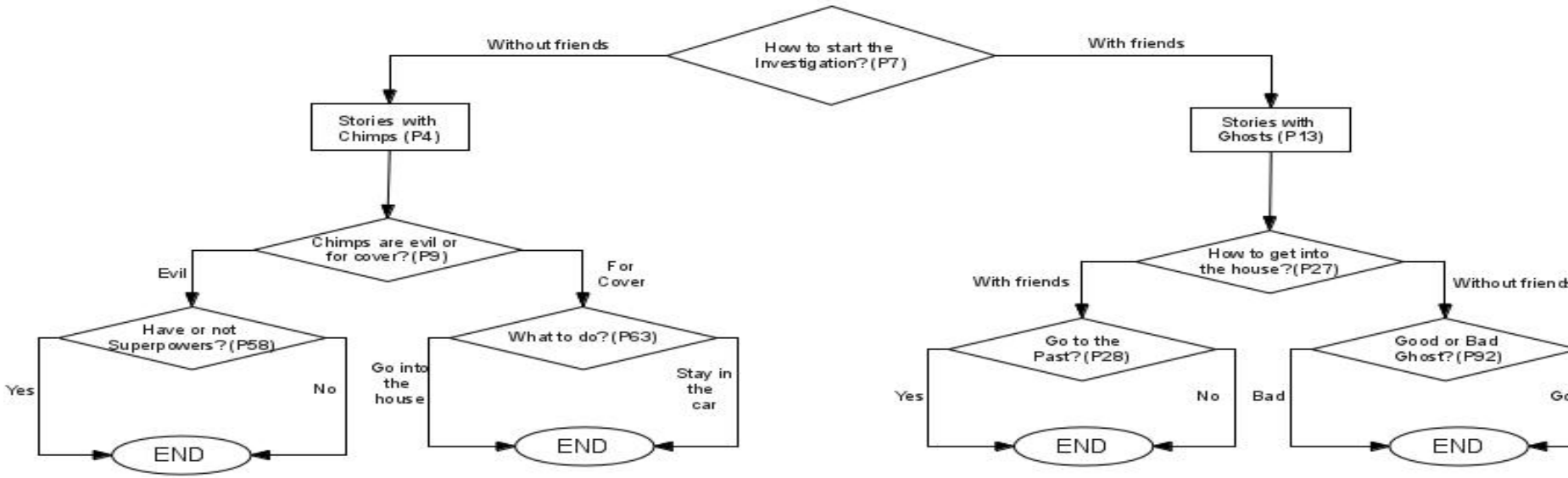
- How many different readings are there?
- *Count/enumerate ``walks''*
- What is the length of each reading? Which readings are shortest/longest?
- *Count shortest path, longest ``walk''*
- Is any storylet repeated in a single reading?
- *Is the graph a DAG? Find cycles*

Maths and Graphs

What can graph and matrix based mathematical models tell us about ergodic literature?

- How many different readings are there?
- *Count/enumerate ``walks''*
- What is the length of each reading? Which readings are shortest/longest?
- *Count shortest path, longest ``walk''*
- Is any storylet repeated in a single reading?
- *Is the graph a DAG? Find cycles*
- *Can we rank storylets (... and why?)*
- *Algebraic graph theory*
- *Link based ranking*

Why? Maybe improve on the digraph



Concept maps, perhaps?

Counting walks

The Analysis of Sociograms using Matrix Algebra
Leon Festinger
Human Relations 1949; 2; 153

The number of walks of length k from node i to node j in the graph is the value of $[A^k]_{i,j}$.

Wikipedia

Leon Festinger

Born	May 8, 1919 New York City
Died	February 11, 1989 New York City
Fields	Psychology
Known for	cognitive dissonance

The SMRank metric for storylets

Basic idea

- Ranking based on the level of storylet participation in all possible plots.
- Define scheme in which the rank of every storylet is determined by the number of plots containing it.

Definition

Let $G = (V, E)$ be a DAG. For every $v_j \in V$ with $n = |V|$ let

$$\tau_j := \#(\text{paths in } G \text{ containing } v_j).$$

Then $\texttt{SMRank} : V \to \mathbb{R}$ is defined by

$$\texttt{SMRank}(v_j) = \frac{\tau_j}{\sum_{j=1}^{n} \tau_j}.$$

Computing SMRank for DAGs

Proposition

Assume that $G = (V, E)$ is a DAG with single source v_1 and f sink nodes, v_1 and $v_{n-f+1}, \ldots, v_n$ numbered last, where $n = |V|$. Then for any $v_j \in V$

$$\tau_j = e_1^\top (I - A)^{-1} e_j e_j^\top (I - A)^{-1} \begin{pmatrix} 0 \\ I_f \end{pmatrix} e^{(f)}. \tag{3}$$

Moreover

$$\texttt{SMRank}(v_j) \leq \texttt{SMRank}(v_1).$$

SMRank for general graphs

Difficulties

- When cycles are present, there can be infinite walks.
- ... counting becomes pointless.
- In matrix terms, $I - A$ becomes singular.

Issues

- discount ``redundancies''
- count judiciously combining paths and cycles
- *combinatorially difficult* (possibly $\#P$)

Analogy: Ergodic reader ∝ Random surfer

reader	**surfer/crawler**
open book	turn browser on
goto starting storylet	open homepage
choose ``next'' storylet	click on existing link
reached ending, goto start storylet	reached sought page, click home
reached ending, stop	reached sought page, stop
choose any storylet	enter URL, click some website



Link-based ranking and PageRank

What makes a web page important for Google?

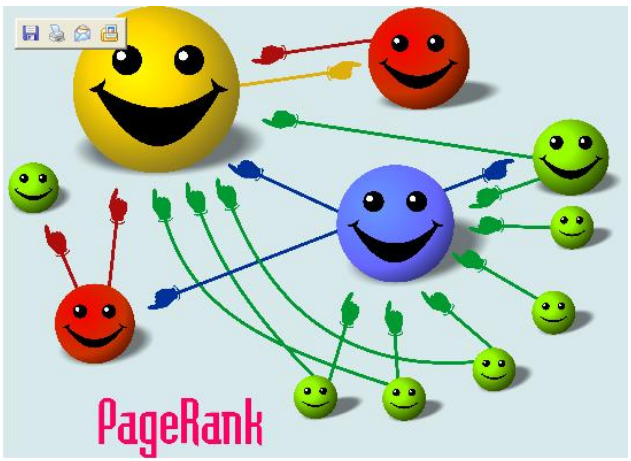
- many important pages contain links to it
- a page containing many links has reduced impact on the importance of the pages it contains links to

Ranking equation: $x_i = \sum_{P_j \in B_i} \frac{x_j)}{|P_j|}$ where x_i is *rank of node* P_i and $B(i)$ the set of incoming links to P_i.

The PageRank vector

$$x^{\mathrm{PR}} := [x_1, \ldots, x_n]^\top$$

``expresses'' the relative importance of webpages.



Anatomy of Google matrices and PageRank vectors

Remarks: PageRank defined for general digraphs ... after some adjustments Motivation: Existence, uniqueness and ease of computation

$$A \overset{\text{de-dangle}}{\rightarrow} \hat{P} = A^\top + wd^\top \overset{\text{stochastize}}{\rightarrow} S = \hat{P}D^{-1}$$

$$\begin{aligned} G(\mu) &= \mu S + (1-\mu)ve^\top \\ x^{\mathrm{PR}} &= G(\mu)x^{\mathrm{PR}} \text{ where } e^\top x^{\mathrm{PR}} = 1 \\ x^{\mathrm{PR}} &= (1-\mu)(I-\mu S)^{-1}v \end{aligned}$$

Remarks

When $\rho(\mu S) < 1$ can expand in Neumann series:

$$(I - \mu S)^{-1} = I + \mu S + \mu^2 S^2 + \cdots$$

μ acts as attenuation factor damping higher order terms.

- Predecessor ranking by Katz (1953) to rank status of individuals.
- Essentially same as PageRank when all nodes have equal outdegree.
- Brin&Page used it to make the power method convergent.

Problem specific observations

Natural irreducibility

- Inclusion of link from every ending storylet to the starting storylet makes them strongly connected and the matrix irreducible.
- Existence of eigenvalue $\rho(\mu) = 1$ and positive eigenvector (Perron-Frobenius theory)

Alternative role for damping parameter

- Matrices of moderate size so no need to use the power method that stalls when matrix is imprimitive. Perron eigenvector can be computed directly.
- Choosing $\mu \in (0, 1)$ permits the random surfer interpretation of PageRank (Brin,Page)
- Parameter chosen to discount the contribution of longer readings.
- Random surfer analogy with ergodic reader.

The CHILDIF collection

12 books, 1500 pages

Choose Your Own Adventure (CYOA corp.) 1979-1998, www.cyoa.com/	>200 (6)
Innerstar University (American Girl) 2010-today, http://web.innerstaru.com/	9 (4)
Multiclone Tales (Kalendis pub.) 1997-2003, http://www.kalendis.gr/	2 (2)

title	name	*n*	nnz
Choose Your Own Adventure			
Abominable Snowman	CYOA_AS	91	93
Journey Under the Sea	CYOA_JU	101	109
Space and Beyond	CYOA_SB	115	119
Lost Jewels of Nabooti	CYOA_LJ	110	114
Mystery of the Maya	CYOA_MM	113	116
House of Danger	CYOA_HD	91	90
Innerstar University			
Girl's Best Friend	INUN_GB	110	116
Taking the Reins	INUN_TR	103	107
Into the Spotlight	INUN_IS	112	115
Fork in the Trail	INUN_FT	110	122
Multiclone Tales			
88 Dolmadakia	MUTA_ED	154	265
33 Pink Rubies	MUTA_TP	211	386

Statistics

for DAG books

title	storylets	ends	plots	lengths (min, max, avg)
CYOA_AS	91 (+25)	28	36	(7,20,11.8)
CYOA_MM	113 (+18)	39	106	(8,30,19.7)
CYOA_HD	91 (+17)	20	20	(9,19,14.4)
INUN_GB	110 (+11)	24	68	(8,24,17.5)
INUN_TR	103 (+19)	24	37	(10,25,16.3)
INUN_IS	112 (+11)	24	40	(7,27,17.8)
INUN_FT	110 (+11)	23	511	(7,34,25.9)

Statistics

for non-DAG books

title	storylets	links	ends	plots
CYOA_JU	101 (+16)	109	42	(>202)
CYOA_SB	115 (+16)	119	43	(>98)
CYOA_LJ	110 (+21)	114	38	(>92)
MUTA_ED	154	265	41	(>1349)
MUTA_TR	211	386	53	(>220431)

Software tools

- MATLAB

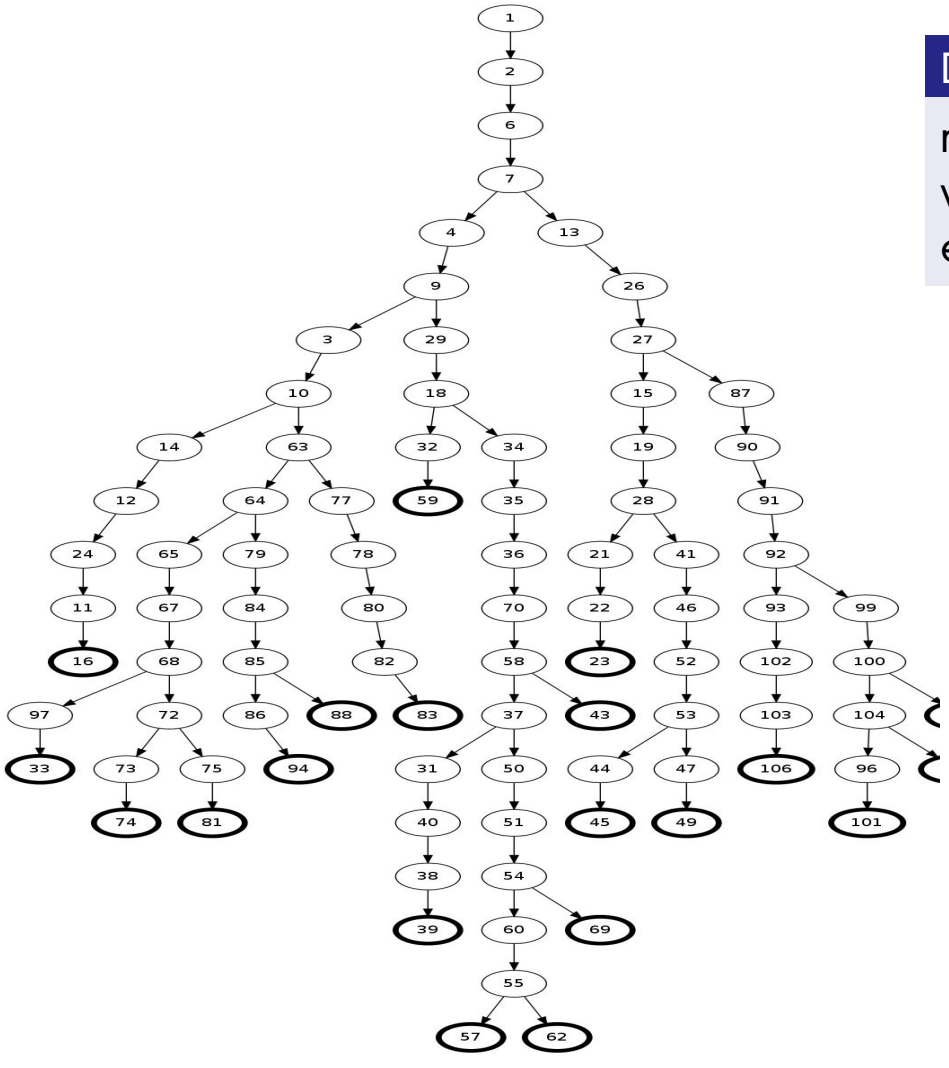
- GraphViz

- MatlabBGL by D. Gleich

- `graph_to_dot.m` by L. Peshkin for GraphViz

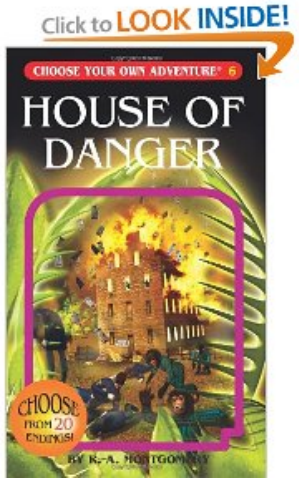
- Brain Connectivity Toolbox (`findpaths.m`) by O. Sporns

House of Danger (CYOA_HD, $n = 91$, $\mathtt{nnz} = 90$)

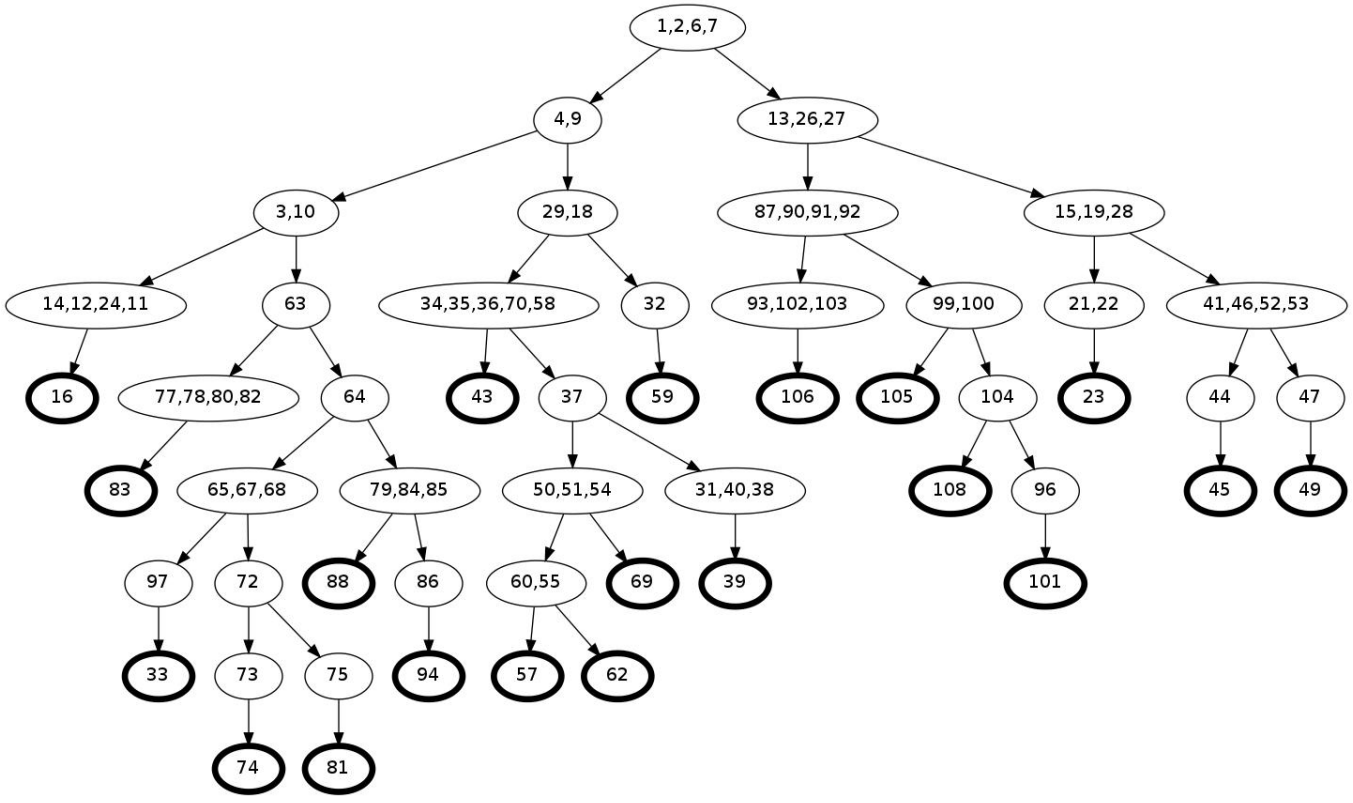


DimRedct

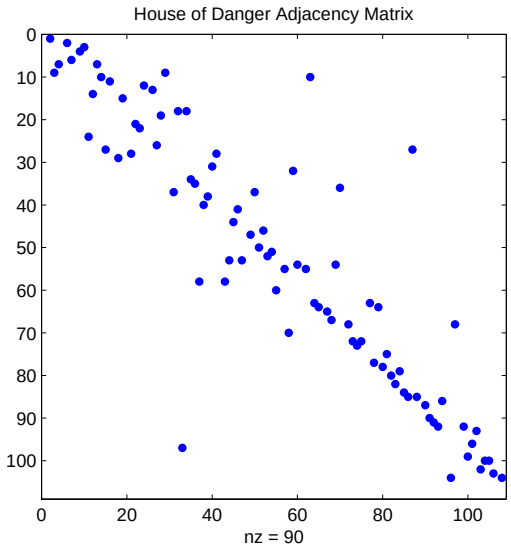
merge node sequences into supernodes with sole branch node at end node of each sequence $\Rightarrow$



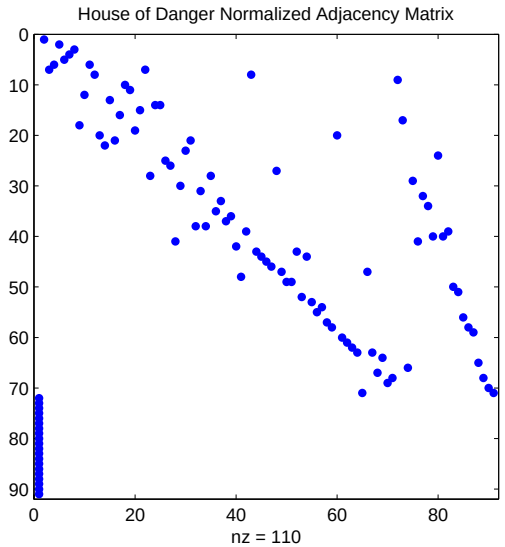
CYOA_HD after dimensionality reduction



spys at the House of Danger



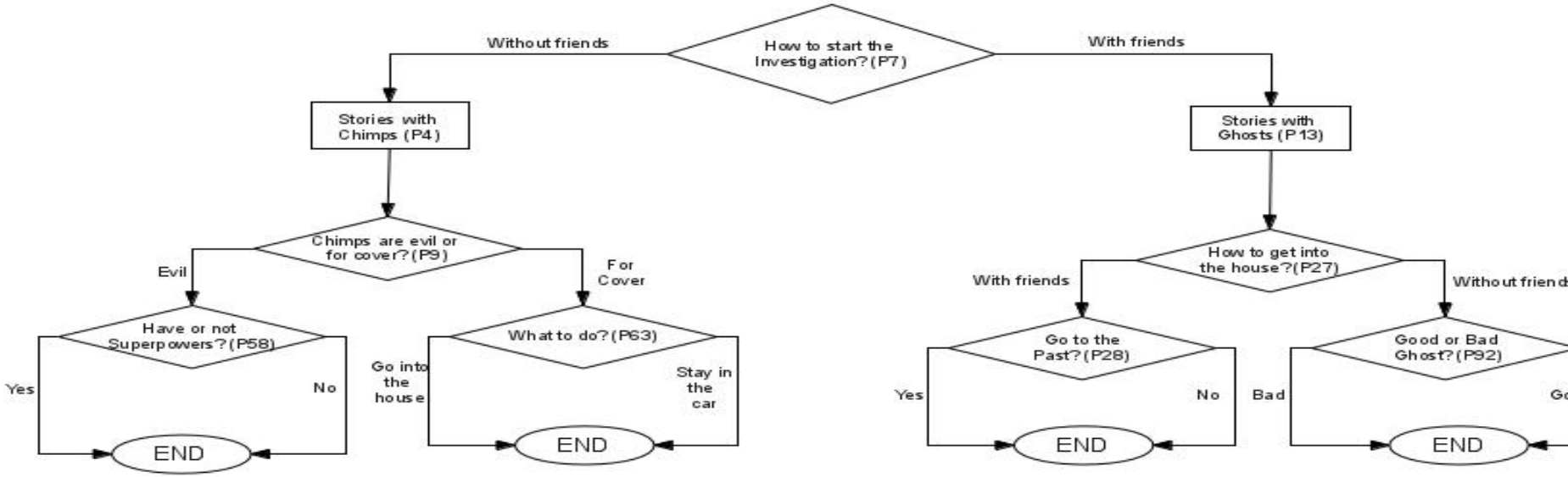
(a) Original



(b) de-dangled

Concept Map for CYOA_HD

Represents main narrative axes around which most plots revolve as discovered by ``independent readers''. CM's provide a view of the book's ```high-level structure''.

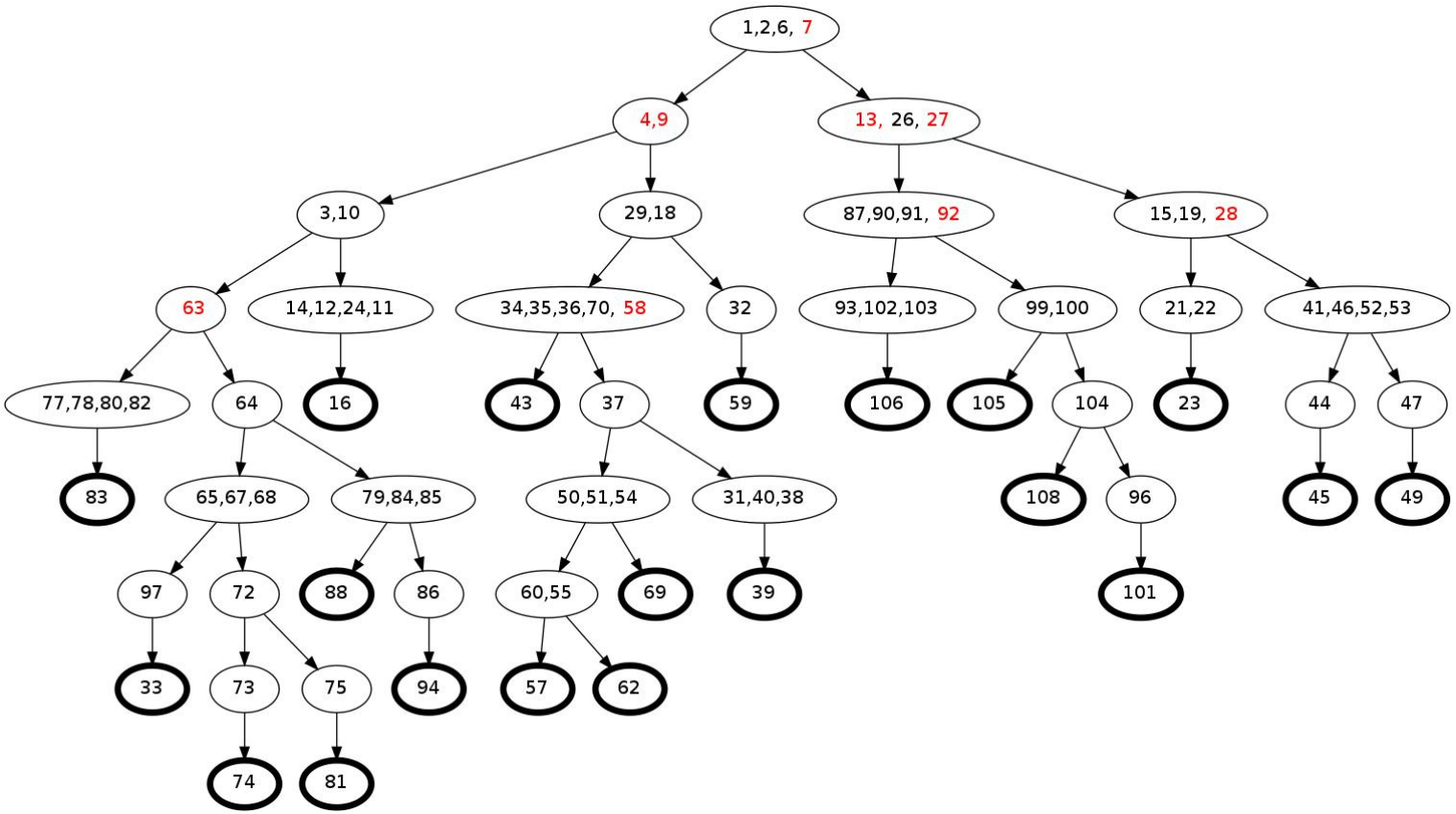


Is there any relation between the CM and the rankings?

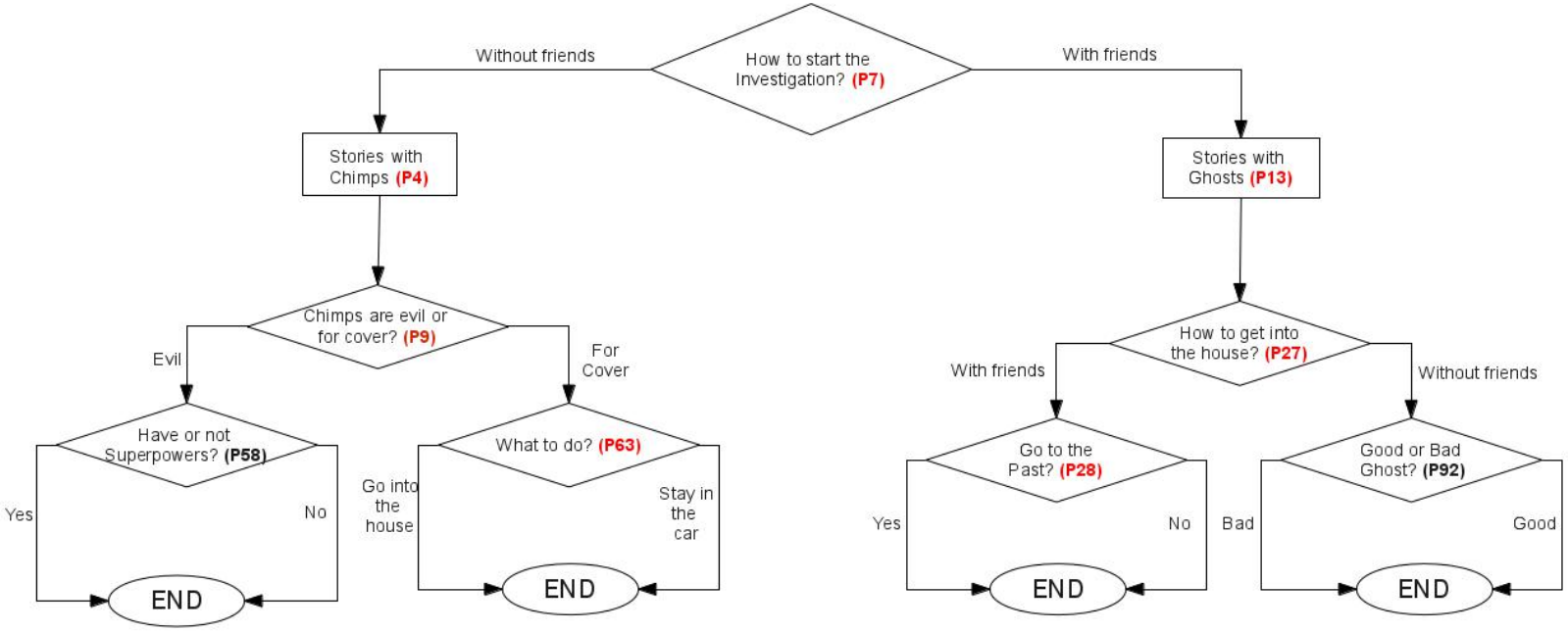
SMrank top scores ($v = \frac{1}{n}e$)

SMRank		$PR(0.85)$		$PR(1.0)$	
rank	node	rank	node	rank	node
0.0652	7	0.0973	1	0.0711	1
	6	0.0844	2		2
	1	0.0734	6		6
	2	0.0640	7		7
		0.0289	4	0.0356	4
0.0404	4		13		13
	9	0.0262	26		9
			9		27
0.0248	13	0.0239	27		26
	26	0.0128	3	0.0178	3
	27		29		28
0.0217	3	0.0125	10		29
	10		18		87
	63	0.0118	15		91
	29		87		92

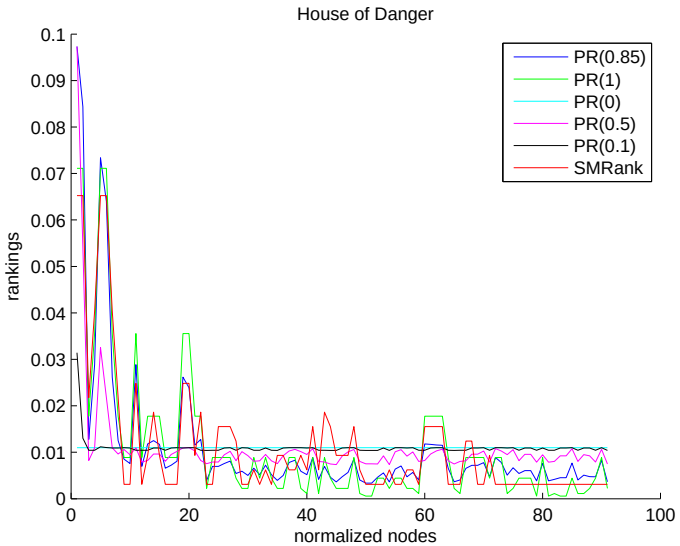
Results on the digraph

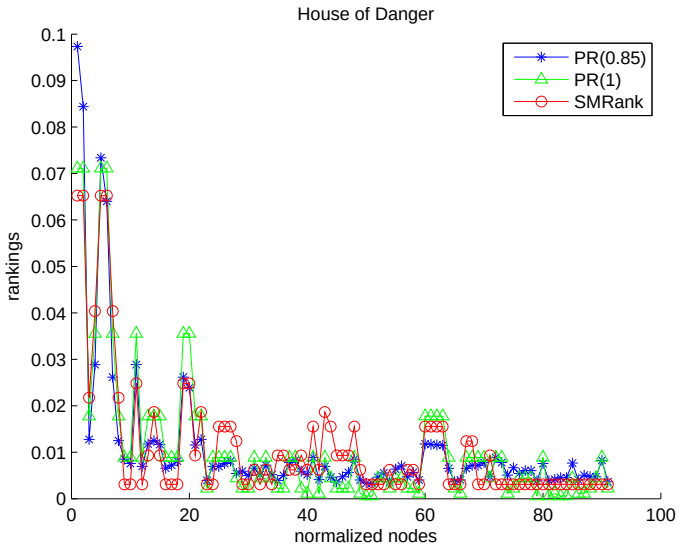


Concept Map for CYOA_HD



Comparing SMRank with PageRank





Conclusions

What can graph and matrix based mathematical models tell us about ergodic literature?

matrix and graph theoretic techniques have the potential to extract latent information from ergodic literature -- hard to obtain by simply reading or graph browsing.

- Created data collection CHILDIF for general use. This motivates the use of graph and Web based metrics for ranking storylets and some other tasks.
- Several metrics expressed and computed with matrix operations.
- `SMrank` provides guidance but cannot be used with cycles
- Random surfer provides a ueful model for reader behavior.
- Link-based ranking like PageRank detects ``important'' plot decisions.

Εργασίες από ACM Hypertext 2012 και ACM Hypertext 2013

Graph and Matrix Metrics to Analyze Ergodic Literature for Children

Eugenia-Maria Kontopoulou
CEID, University of Patras
Rio, Greece
kontopoulo@ceid.upatras.gr

Maria Predari
CEID, University of Patras
Rio, Greece
predari@ceid.upatras.gr

Thymios Kostakis
SEOKO
28, A. Papandreou st.
Halandri, Athens, Greece
kostakth@gmail.com

Efstratios Gallopoulos
CEID, University of Patras
Rio, Greece
stratis@ceid.upatras.gr

Onomatology and content analysis of ergodic literature

Eugenia-Maria Kontopoulou
CEID, University of Patras
Rio, Greece
kontopoulo@ceid.upatras.gr

Maria Predari
CEID, University of Patras
Rio, Greece
predari@ceid.upatras.gr

Efstratios Gallopoulos
CEID, University of Patras
Rio, Greece
stratis@ceid.upatras.gr

Τέλος Ενότητας

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 5: Charles Babbage 1791-1871

http://www.don-lindsay-archive.org/talk/babbage.html

Διαφάνεια 5: Ada Byron Lovelace 1815- 1852

https://en.wikipedia.org/wiki/Ada_Lovelace

Διαφάνεια 6: Analytical engine

https://en.wikipedia.org/wiki/Analytical_Engine#/media/File:Analytical_Engine_(2290032530).jpg

Διαφάνεια 10: Patriot misile failure, 1991

https://www.ima.umn.edu/~arnold/disasters/patriot.html

Διαφάνεια 10: Explosion of Ariane 5, 1996

https://www.ima.umn.edu/~arnold/disasters/ariane.html

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 10: Sinking of Sleipner A offshore platform, 1991

http://www.ima.umn.edu/~arnold/disasters/sleipner.html

Διαφάνεια 17: Competing on analytics

http://www.amazon.com/Competing-Analytics-The-Science-Winning/dp/1422103323

Διαφάνεια 18: Netflix Prize

http://www.weigend.com/files/teaching/stanford/2008/stanford2008.wikispaces.com/6.html

Διαφάνεια 19: Netflix Prize

http://www.netflixprize.com/

Διαφάνεια 21: Diels 1848- 1922

http://www.ulb.ac.be/assoc/aip/diels_150dpi.jpeg

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 21: Antike technik

http://www.weiss-leipzig.de/teubner-hall-of-fame-1845-bis-2000.htm

Διαφάνεια 21: Θησαυρός της ελληνικής γλώσσας (TLG)

http://www.tlg.uci.edu/images/TLGdisk.jpg

Διαφάνεια 37: Undestanding complex data sets

https://www.crcpress.com/Understanding-Complex-Datasets-Data-Mining-with-Matrix-Decompositions/Skillicorn/9781584888321

Διαφάνεια 53: Examples of networks- The WWW

http://lev.ccny.cuny.edu/~hmakse/complexnets.html

Διαφάνεια 54: Examples of networks- Obesity

https://thesituationist.wordpress.com/2008/03/03/social-networks/

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 55: L. Page, S. Brin

http://mooglemb.com/threads/googlecache/64.233.187.104/corporate/execs.html

Διαφάνεια 55: The Earth Simulator Center

https://arctelix.wordpress.com/about/earth-simulator-center/

Διαφάνεια 57, 92: PageRank

https://en.wikipedia.org/wiki/PageRank

Διαφάνεια 58: Η μέθοδος PageRank της Google και άλλα συστήματα κατάταξης

http://www.cup.gr/Previews/978-960-524-313-5-Preview.pdf

Διαφάνεια 69, 77: The 88 dolmadakia

http://www.goodreads.com/book/show/15938307-88

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 75-76: Hopscotch

http://www.goodreads.com/book/show/53413.Hopscotch

Διαφάνεια 75-76: Afternoon, a story

http://www.eastgate.com/catalog/Afternoon.html

Διαφάνεια 75: Hypertext 3.0

http://www.amazon.com/Hypertext-3-0-Critical-Globalization-Re-visions/dp/0801882575

Διαφάνεια 75-76: A heartbreaking work of staggering genius

https://en.wikipedia.org/wiki/A_Heartbreaking_Work_of_Staggering_Genius

Διαφάνεια 75-76: Παράφρον κρέας

http://www.kapsimi.gr/parafron-kreas

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 77: Τα 33 ροζ ρουμπίνια

https://www.ianos.gr/ta-33-roz-roumpinia-0022858.html

Διαφάνεια 77: Sugarcane Island

https://en.wikipedia.org/wiki/Edward_Packard

Διαφάνεια 77: A Girl's Best Friend

http://americangirl.wikia.com/wiki/A_Girl%27s_Best_Friend

Διαφάνεια 77: Space and Beyond

http://www.amazon.com/Space-Beyond-Choose-Your-Adventure/dp/1933390034

Διαφάνεια 77, 101: House of Danger

http://www.amazon.com/House-Danger-Choose-Your-Adventure/dp/1933390069

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 77: Fork in the Trail

http://americangirl.wikia.com/wiki/Fork_in_the_Trail

Διαφάνεια 77: Mystery of the Maya

http://www.gamebooks.org/show_item.php?id=555

Διαφάνεια 79: Aarseth

http://forskning.no/spill-media-internett-medievitenskap-samfunnsokonomi-sosiologi-stub-data/2008/02/dataspill-viser

Διαφάνεια 79: Cybertext

https://books.google.gr/books/about/Cybertext.html?id=qx_-zj0-TwoC&redir_esc=y

Διαφάνεια 81: Raymont Queneau

http://www.nndb.com/people/769/000113430/

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Διαφάνεια 87: Leon Festinger

https://en.wikipedia.org/wiki/Leon_Festinger#/

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει) μαζί με τους συνοδευόμενους υπερσυνδέσμους.