Nola aurkitu orratza lastategian

2013-2014 ikasturtearen
hasiera-hitzaldia

Luis M. Ezquerro
irakasle doktore jauna

Nafarroako Unibertsitate Publikoko
Aljebrako katedradunak

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

upna

Universidad
Pública de Navarra
Nafarroako
Unibertsitate Publikoa

Nola aurkitu orratza lastategian

2013-2014 ikasturtearen
hasiera-hitzaldia
Luis M. Ezquerro
irakasle doktore jauna

Nafarroako Unibertsitate Publikoko
Aljebrako katedradunak

Iruñean, 2013ko irailaren 13an

upna
Universidad
Pública de Navarra
Nafarroako
Unibertsitate Publikoa

Argitaratzailea: Nafarroako Unibertsitate Publikoa

Koordinazioa: Komunikazio Zerbitzua

Fotokonposizioa: Pretexto. pretexto@pretexto.es

Inprimatzailea: Ona Industria Gráfica

Lege gordailua: NA 1286-2013

Banaketa: Argitalpen Atala
Nafarroako Unibertsitate Publikoa
Arrosadiko campusa
31006 Iruñea
Faxa: 948 169 300
Posta: publicaciones@unavarra.es

Erektoreak 2013-14 ikasturteko hasierako hitzaldi hau egiteko eskatu zidanean, adierazteko zailak diren sentimendu ugari metatu ziren nire baitan: esker ona zeregin garrantzitsu hori egiteko ardura eman didanari, batzuetan beren burua sakrifikatuz kantore honetara iristeko ibilbidea erraztu zidatenen eta didatenen oroitzapena, urteen joan geldiezinaren kontzientzia, ohorea, ardura eta poza nire kideei, eta, oro har, hemen ordezkatua dagoen gizarteari nire eguneroko zereginaren ezaugarri batzuk azaltzeagatik.

Hitzaldia egiteko mandatua jaso nuen egun haietan Informatika Ingeniaritzako Graduko nire ikasleei Grafoen Teoriari buruzko atariko nozio batzuk azaltzen ari nintzen. Artean esaten ari nintzaien 2013. urte honetan betetzen direla ehun urte Paul Erdös matematikari ospetsua Budapesten jaio zela. Abiapuntu ona iruditzen zait zuei nahi dudana kontatzeko. Erdösek egundoko ekarpena egin zion Matematikari. Hainbat gairekin lotutako kontuetan lan egin zuen. Esaterako, konbinatoria, grafoen teoria, zenbakien teoria, analisi klasikoa, multzoen teoria, probabilitateen teoria... Zalantzarik gabe, garai guztietako matematikari emankorrenetakoa da.

Paul Erdös izaera bereziko zientzialaria izan zen, baten batek bere gisakoa esango luke, eta berea dela diote definizio bitxi hau: matematikaria da kafeak teorema bihurtzen dituen makina. Jainkoaren existentziaz zalantza egiten bazuen ere, eta hari «Faxista Gorena» deitzen zion, esaten zuen Jainkoak LIBURUA idatzi zuela eta irudizko liburu horretan erregistratu zituela teorema guztiak beren froga ederrenekin eta argigarrienekin; berak zioenez, *ez da beharrezkoa Jainkoagan sinestea, baina ezinbestekoa da LIBURUAN sinestea*.

Erdösen LIBURUA jakintza unibertsal bat jasotzen duen irudizko artxiboaren adibide bat baizik ez da. Eta Literaturak eskaintzen dizkigu liburutegi unibertsalak protagonista dituzten fikzioen adibideak. Iraganeko, gaur egungo eta etorkizuneko liburu guztiak dituen artxibo baten ideia kitzikagarriari gehitzen zaio liburu kopuruak sortzen duen harridura, mugatua izanagatik ere, gizakiaren ulermenetik harago baitago. *Nola izan daiteke mugatua agortezina dena?*[1] galdetzen dio bere buruari Kurd Laßwitz idazle alemanak 1904an argitaratu zuen «Liburutegi unibertsala ipuinean»[2]. Liburutegi horretako liburuak bi zentimetro lodi izango balira eta ilaran jarriko balira, argi izpi batek $2 \times 10^{1.999.982}$ urte inguru beharko lituzke liburutegia zeharkatzeko. *Ezin da ulertu* –dio– *ez zenbat urte behar dituen argiak liburutegia zeharkatzeko, ez zenbat liburuki diren ere*[3]; pentsa dezagun unibertsoak 14 mila milioi urte dituela jotzen dugula, hau da, $1{,}4 \times 10^{10}$ urte inguru.

Berrogei urte baino gutxiago igaro ziren Jorge Luis Borgesek «El jardín de los senderos que se bifurcan» izenburuko kontakizun laburren bilduma aurkeztu zuen arte. Bilduma horretan sartu zuen *La biblioteca de Babel* ospetsua, eta horren ideia Laßwitz-en neurririk ezatik hartu zuen[4]. Borgesen liburutegian ez daude soilik Tazitoren lan galduak, Laßwitz-enean bezala, baizik eta *etorkizunaren historia zehatza*, eta, gainera, *zure heriotzaren egiazko kontakizuna edo goiaingeruen autobiograak*. Eta alemaniarrarenean bi adiskide garagardo edaleren arteko elkarrizketa atsegina zena, beharbada hain onuragarria den edariaren ondorio, argentinarrarenean ikuspegi kezkagarri, nahasgarri eta labirintikoa da.

Itxurazko amaigabetasun horretan liburu jakin bat nola aurkitu behar den aztertzen dute bai Laßwitzek bai Borgesek. Biak bat datoz honetan: liburua aurkitzeko probabilitatea, Borgesen hitzetan, *zeron konputagarria* da. Gaztelaniaz bada esaera zahar bat, zailtasun praktiko ia gaindiezin hori modu esanguratsuan azaltzen duena:

1. Wie soll das Unerschöpfliche endlich sein?

2. Laßwitz: *Die Universalbibliothek*. 1904ko abenduaren 18an argitaratua Breslauko *Ostdeutschen Allgemeinen Zeitung* de Breslau. <http://gutenberg.spiegel.de/buch/3130/1>. Gaztelaniazko bertsioa: A. Hanke-Schaefer eta A. Fernández Ferrer, helbide honetan: <http://www.letraslibres.com/revista/convivio/borges-y-sus-precursores>

3. Man kann sich die Zahl der Jahre, die das Licht braucht, an der Bibliothek entlangzulaufen, ebensowenig vorstellen, wie die Zahl der Bände selbst.

4. J. L. Borges: *La biblioteca de Babel*, liburu honetan: «El jardín de los senderos que se bifurcan» (1941). *Obras Completas*. Círculo de Lectores, 1992.

lastategian orratza bilatzea da hori. Eta hala, alemaniarrak, eragozpen honen aurrean, baikortasunez baieztatzen du:

> «Zentzumenei dagokiena, denborarekin, iragankorra da. Logikoa denak ez dauka loturarik denborarekin eta unibertsala da. Eta logikoa dena gizateriaren pentsamendua besterik ez denez, horregatik baitugu dohain denboragabe hau, zeinari esker partekatzen baititugu jainkozkoa denaren lege betierekoak, berdin partekatzen dugu ahalmen sortzaile mugagabearen patua ere bai. Horretan datza Matematikaren oinarrizko legea»[5].

baina argentinarrak honela dio:

> «Neurrigabeko itxaropenaren ondoren, espero izatekoa denez, gehiegizko depresioa etorri zen. Ia jasangaitza zirudien ziur jakiteak hexagonoren bateko apalen batean liburu zoragarriak zeudela, eta liburu zoragarri horiek eskuraezinak zirela»[6].

Laßwitz 1910ean hil zen; Borges, 1986an. Ez dakigu zer pentsatuko zuten biek gure garaietako artxiboen artxiboa ezagutu izan balute, ingeles hiztunek, zentzuz, *web* deitzen duten hori, hau da, amarauna. Idazleen irudimenak sortu zituzten liburutegiak baino txikiagoa bada ere, web sarea iritsi da dagoeneko Laßwitzek *agortezin* deitzen duen horren mugetara. *web*-ak ere aurre egin behar izan zion, behar zuen garaian, bilatzaileen auziari, baina, literaturako aurrekoetan gertatzen ez zen bezala, emaitzak gaur harri eta zur uzten gaitu bere arintasunagatik eta eraginkortasunagatik. Edozein matematikari, eta, bereziki, edozein aljebralari harro dago jakiteagatik Aljebra Linealaren teorema sorta bat dela *orratza lastategian* bilatzeko prozedura bikain hauen sekretua. Saiatuko naiz hitzaldi honetan azaltzen, zirriborratuz, zein den webean dauden bilatzaileen oinarri algebraikoa.

1. Zer gertatzen da *web*-ean bilaketa bat egiten denean?

Bilatzaileen gaia zabala eta askotarikoa da; espezialista dokumentatuenen esku utziko dut konputazioaren mundu formaniztunari dagokion guztia. Bilatzaile baten funtzionamenduaren muina osatzen duten matematika kontuetan oinarrituko badut

5. Das Sinnliche ist vergänglich mit der Zeit, das Logische ist unabhängig von aller Zeit, ist allgemeingültig. Und weil dieses Logische nichts anderes bedeutet als das Denken der Menschheit selbst, so haben wir in diesem zeitlosen Gut einen Anteil an den unwandelba- ren Gesetzen des Göttlichen, an der Bestimmung der unendlichen Schöpfermacht. Darauf beruht das Grundrecht der Mathematik.

6. A la desaforada esperanza, sucedió, como es natural, una depresión excesiva. La certidumbre de que algún anaquel en algún hexágono encerraba libros preciosos y de que esos libros preciosos eran inaccesibles, pareció casi intolerable.

ere nire hitzaldia, iruditzen zait merezi duela azaltzea, labur bada ere, aldez aurretiko xehetasun batzuk.

Nabarmendu behar den lehendabiziko gauza hau da: konpainien bilatzaileek ez dute beren lana *web* osoan egiten, baizik eta sarearen indize batean, edo gutxienez konpainiari sartzeko aukera ematen dion eta sarearen parte den indize batean. Bilatzaile baten aurreneko zeregina da *armiarma* izeneko robot batzuk bidaltzea (ez dezagun ahaztu «amaraun» batean gaudela). Robot horiek *web* orrialdeen multzo batzuk aztertzen dituzte, dauzkaten estekei jarraitzen diete, eta beren lana errepikatzen dute[7]. Armiarmek lortutako informazioa biltegi nagusi batean gordetzen da behin-behinean. Indexatze moduluak datu batzuk atera, konprimatu, eta zenbait indizetan gordetzen ditu.

Agerikoa da erabiltzaileak giza hizkuntza erabiltzen duela. Kontsulta bateko testua eraldatuko ez balitz, emaitzak eskas urriak lirateke, edo, beharbada, hutsaren hurrengoak. Eraldaketa horren helburua da bilatzaileak uler dezala zer eskatzen diogun, eta eskatzen diogunari dagokion informazio kopuru handiena emanez erantzun diezagula. Horregatik desagertzen dira testutik tildeak, letra larriak eta bilaketaren zentzua funtsean aldatzen ez duten hitz guztiak: preposizioak, artikuluak, izenordainak, juntagailuak... Hitz horiek *stopword*[8] dute izen teknikoa. Halaber, *lematizazio*[9] izeneko hizkuntza prozesu bat egiten da, hau da, hitz baten erroa atera, eta hitzaren forma jokatu guztien ordezkotzat hartzen da

Eraldatze prozesu horiek burutu eta gero, kontsultara egokitzen diren orrialdeak bilatzen ditu softwareak. Ziur asko, termino horiek dauzkaten ehunka mila orrialde aurkituko ditu, baita milioiak ere.

Hitzez hitzeko edozein bilaketak bi arazo dauzka: polisemiarena eta sinonimiarena. Nola gainditu arazo horiek? Zalantzarik gabe, hitzez hitzeko bilaketa egin beharrean, esanahiaren araberako bilaketa egin behar dugu. Aljebra Linealak bektore-espazioaren eredua ematen digu. Sinplikatuz, dokumentu bakoitza bektore bat balitz bezala kodetzen da, eta bektore horren koordenatu bakoitzak, osagai bakoitzak, erakutsi egiten du nolako garrantzia duen termino jakin batek, kontzeptu batek edo gako-hitz batek dokumentu baten semantikan. Kontzeptuak dokumentuan edo dokumentu bilduman duen maiztasunaren funtzioa izaten da balio esleitua edo ponderazioa, baina kontuan hartzen du, halaber, non eta nola agertzen den terminoa: izenburuan, testu idatzian

7. Bilaketa prozesuan egiten den aurreneko lan horri *web crawling* esaten zaio ingelesez.

8. Esteka honetan dituzue gaztelaniaz balio duten ia 400 *stopword*: <http://droope.org/2011/02/28/stopwords-para-espanol-castellano/>.

9. Ingelesez *stemming* esaten zaio.

letra nabarmenagoz (letra lodietan, etzanetan, maiuskula txikietan, azpimarratuetan), beste orrialde batzuetarako estekak dituzten testuetan, etab.

Bektore mintzaira honen bidez, kontsulta kodetzen duen bektoretik hurbilen dauden bektoreen detekzio bihurtzen da kontsulta zehatz bati dagozkion dokumentuak aurkitzeko ideia. Azkenean, bektore horiek osatzen duten angeluaren anplitudea neurtzea da kontua. Aljebra Linealak gure esku jartzen duen ohiko tresna da biderkadura eskalarra. Hala ere, sinonimiak eta polisemiak sorturiko «zarata» hain handia da, non metodo hau landugabe samarra baita. Hobetu egin behar da, eta horretarako nola ez, ba! Aljebra Linealaren beste emaitza askoz ere landuago batera jotzen da: matrize baten *Balio Singularren Deskonposizioa*.

Xehetasunetan sartu gabe, esan dezagun hemen, edozein tamainatako matrize baten balio singularren deskonposizioak eskaintzen digun eraikuntzari esker, informazio galera handirik gabe iragazten dela aipatzen genuen «zarata» hori.

Aljebrako teknika horien bidez dokumentuen zerrenda ordenatu bat lortzen dugu. Baina gure kontsultak baldintzatzen du ordenazio hori, eta ez du kontuan hartzen dokumentu bakoitzaren garrantzia. Ez dira, ordea, garrantzi berekoak dokumentu guztiak, eta batzuk nahitaez aipatu beharrekoak badira esparru batean, beste batzuek tokian tokiko eragin mugatua besterik ez dute. Garrantziaren alde horrek gure kontsultarekin loturarik ez duen pisu bat ematen du, aurreko azalpenean kontuan hartu ez duguna.

Baina nola neurtzen dugu dokumentu baten garrantzia? Zientzia ingurunetan ohitu gara dagoeneko eraginaren zenbait parametro eta adierazle erabiltzera. Horien bidez saiatzen gara neurtzen zientzia-aldizkarien kalitatea, eta, funtsean, aipamen kopuruan oinarritzen dira. Haien erabilerak, eta batez ere gehiegizko erabilerak, eztabaida ugari sortu du.

Eredu horietatik ideia hartuz agertu zen *PageRank web* orrialdeei garrantzia esleitzeko algoritmoa. Algoritmoaren ezarpena eta garapena izan dira bilatzaile informatikoek duten boterearen erakusle txundigarrienetako bat: Google[10].

Guri dagokigunez, *PageRank* ereduaren oinarria da buru argiz erabiltzea XX. mendearen hasieratik ezagutzen diren Aljebra Linealeko teorema erabakigarri batzuk: Perron-Frobeniusen teoremak.

Hitzaldiaren gainerako ataletan Google bilatzailearen *PageRank* ereduaren matematika oinarriak azalduko ditugu laburki.

10. Bitxikeria gisa esango dugu Google izena *googol* izenetik datorrela. Norbaitek, Edward Kasner matematikariaren iloba batek, antza, izen hori jarri zion 10^{100} zenbakiari.

2. Nola neurtzen du Google bilatzaileak dokumentu baten garrantzia *web*-ean?

Bilaketa bat egiten dugun aldiro, Google-k hiru gauza jakinarazten dizkigu: aurkitutako orrialde kopurua osotara (kopuru oso handia izaten da), zenbat denbora behar izan duen bilaketa egiteko (berehala egiten du), eta aurkitu dituen hasierako *web* orrialdeen zerrenda ordenatu bat, betiere gure kontsultarekin loturaren bat duten orrialdeak badira. Nola izan daiteke hori?

Sergey Brin (2010)

Lawrence Page (2009)

Bilaketa eta ordenamendu algoritmoan dago sekretua. Sergey Brin matematikariak eta Lawrence Page informatikariak diseinatu zuten metodo hori 1998an Stanfordeko unibertsitatean, biak Informatika doktoretzako ikasleak zirenean[11]. Beren algoritmoari *PageRank* deitu zioten, bere sortzaileetako baten abizena aipatzeko, dirudienez.

11. S. Brin eta L. Page: *The anatomy of a large-scale hypertextual web search engine*. Technical Report. Stanford. In: Seventh International World-*Wide web* Conference (WWW 1998), April 14-18, 1998, Brisbane, Australia. <http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf>.

Bada beste dokumentu bat, geroxeagokoa, aurrekoa zabaltzen eta zuzentzen duena: L. Page, S. Brin, R. Motwani eta T. Winograd-en artikulua da. Hau du izenburua: *The PageRank citation ranking: Bringing order to the web*. Technical report, Stanford InfoLab. 1999. <http://dbpubs.stanford.edu/pub/1999-66>.

Zehatz dezagun gure helburua. Interneteko bilatzaileari, eta bereziki Google-ri, kontsulta bat egiten diogunean, gure kontsultari dagozkion web orrialdeen zerrenda bat ematea nahi dugu, eta zerrenda hori garrantziaren arabera ordenatua egotea[12].

Aljebra Linealean datuen zerrenda ordenatu bat bektorearen sinonimo bat da. *P web* orrialde bakoitzari bere garrantziaren araberako neurri bat esleitzen zaio. Neurri horri *P-ren garrantzia* deitzen diogu *I(P)*. Bektore bat bilatzen ari gara, eta *garrantzien bektore* izena emango diogu[13].

$$\begin{array}{lcccc} \text{Orrialde multzoa:} & \{P_1, & P_2, & \cdots, & P_n\} \\ & \downarrow & \downarrow & & \downarrow \\ I\text{: garrantzien bektorea:} & (I(P_1), & I(P_2), & \cdots, & I(P_n))' \end{array}$$

Nola aurkitzen da *I(P)*? Artikulu zientifikoei eginiko aipamenetatik abiatuz eragin indizeak nola egiten diren adibidetzat hartu, eta web orrialde batek daukan garrantziaren aurreneko ideia, ez oso zehatza, halere, emango digu jakiteak zenbat esteka hartzen dituen beste orrialde batzuetatik. Dena dela, gerta daiteke orrialde gutxi batzuetan aipatzea orrialde bat, baina orrialde horiek entzutetsuak izatea, oso aipatuak: «oso garrantzitsuak». Beraz, badirudi garrantzi handiagoa eman behar diegula ez soilik orrialde aipatuenei, baizik eta «orrialde oso garrantzitsuetan» aipatutako orrialdeei ere bai. Aipatuko dugu, bada, *PageRank*-en postulatu nagusia:

PageRank **Postulatua**: Web-*orrialde baten garrantzia proportzionala da orrialde horrekin estekatu diren orrialdeen baturarekin*.

12. Ohar hauek prestatzeko, *web*-ean nabigatzen ari ginela, aipatzea merezi duten dokumentu batzuk aurkitu ditugu. Lehenik eta behin, P. Fernández Gallardoren artikulu hau: *El secreto de Google y el álgebra lineal*. Bol. Soc. Esp. Mat. Apl. 30 (2004) 115-141; halaber, aipatzekoa da J. M. Graciaren aurkezpena: *Álgebra Lineal tras los buscadores de Internet*, Univ. País Vasco, Dpto. Mat. Aplicada, 2002. <http://www.vc.ehu.es/campus/ centros/farmacia/deptos-f/depme/profesor/gracia/buscap.pdf>. Aipatzekoa da, halaber, J. Gimbertek katalanez idatzitako lan hau J. Gimbert: *Les matemàtiques de GOOGLE: l'algorisme PageRank*, Butll. Soc. Cat. Mat., Vol. 26, 1 (2011) 2955. DOI: 10.2436/20.2002.01.33. Azkenik, ez dugu aipatu gabe utzi nahi aurkezpen atsegin bat, ingelesez idatzia, eta ideia-iturri izan duguna, beste gauza batzuen artean, hitzaldi honen izenburua aukeratzeko; *How Google Finds Your Needle in the web's Haystack*, izenburua duen testuaz gari gara. D. Austin da egilea, eta hemen agertu zen: Amer. Math. Soc. Feature Column Monthly essays on mathematical topics. <http://www.ams.org/samplings/feature-column/fcarc-page rank>.

13. Bektoreak zutabe batean baleude bezala erabiliko ditugu, baina, testu hau errazago idazteko, errenkadetan idatzi dira, *transposizioa* adierazten duen prima bat alboan dutela, errenkadak zutabe bihurtzen direla adierazteko, eta alderantziz.

PageRank postulatuaren aplikazio bakunak ekuazio linealen honako sistema hau ematen digu, zeinetan K proportzionaltasun konstantea baita, eta positibotzat joko dugu:

$$I(P_i) = K \left[\sum_{P_j\text{-}k \text{ esteka bat du } P_i\text{-}ra} I(P_j) \right] \qquad (K > 0) \qquad (1)$$

Baina, erne!: postulatu honek akats larri bat dauka: autoerrepikaria da; definitutakoa ez da agertu behar definizioan.

Lehentasunezko kontua da arazo hau konpontzea.

2.1. *Web* sareari lotzen zaion grafo zuzendua

Internet-eko orrialde publikoen multzoak,«web-orrialde» izena duten horien multzoak, grafo zuzenduaren egitura dauka. Hauek ditu osagaiak:

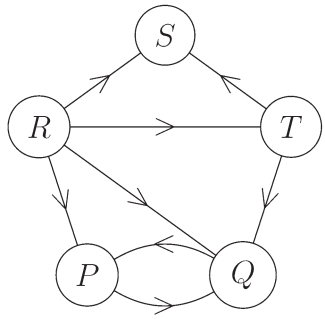
- *Erpin* multzo bat, zeinaren elementuak *web* orrialdeak baitira, eta orrialde bakoitza erpin bat;
- *Gezi* edo *ertz* zuzenduen multzo bat: P orrialdean esteka bat baldin badago Q orrialdera, gezi bat jarriko dugu *P-tik Q-raino*.

Nola bihurtuko dugu izaki grafiko hori kalkulurako egokia den objektu? P_1..., P_n erpinak ordenatu, eta, bakoitzari errenkada bat eta zutabe bat esleituko dizkiogu matrize batean. Kontua da P_j erpinaren zutabean, P_i erpinari dagokion errenkadara iristean, bat zenbakia idaztea, baldin eta P_j*-tik* P_i*-raino* doan gezi bat badago, eta zero zenbakia idaztea ez badago. Grafo zuzenduaren *albokotasun matrizea* izenekoa eraiki dugu.

$$\begin{array}{cc} & \begin{array}{ccccc} P_1 & \cdots & P_j & \cdots & P_n \\ \downarrow & & \downarrow & & \downarrow \end{array} \\ \begin{array}{c} P_1 \rightarrow \\ \vdots \\ P_i \rightarrow \\ \vdots \\ P_n \rightarrow \end{array} & \begin{pmatrix} \vdots & \cdots & \vdots & \cdots & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & \cdots & b_{ij} & \cdots & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & \cdots & \vdots & \cdots & \vdots \end{pmatrix} \end{array} \quad b_{ij} = \begin{cases} 1 & \text{lotura bat baldin badago } P_j\text{-}tik\ P_i\text{-}raino, \\ 0 & \text{loturarik ez baldin badago } P_j\text{-}tik\ P_i\text{-}raino. \end{cases}$$

Demagun, adibidez, *R; S; T; P; Q* bost orrialdeko sare txiki-txiki bat dugula, eta horretan:

- *P* orrialdeak *Q* orrialdea aipatzen duela.
- *Q* orrialdeak P orrialdea aipatzen duela.
- *R* orrialdeak *P; Q; S; T* orrialdeak aipatzen dituela.
- *S* orrialdeak ez duela orrialderik aipatzen.
- *T* orrialdeak *S; Q* orrialdeak aipatzen dituela.



Honako hau da grafo zuzendu horren albokotasun matrizea:

$$M = \begin{array}{cc} & \begin{array}{ccccc} R & S & T & P & Q \end{array} \\ \begin{array}{c} R \\ S \\ T \\ P \\ Q \end{array} & \left(\begin{array}{ccc|cc} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{array}\right) \end{array} = \left(\begin{array}{c|c} M_{11} & (0) \\ \hline M_{12} & M_{22} \end{array}\right)$$

Ez da zaila bost orrialdeen garrantzia kontuan hartu behar duen ekuazio-sistema *PageRank* postulatuaren arabera formulatzea, eta matrize baten bidez adieraztea:

$$\begin{cases} I(R) = 0 \\ I(S) = K(I(R) + I(T)) \\ I(T) = K \times I(R) \\ I(P) = K(I(R) + I(Q)) \\ I(Q) = K(I(R) + I(T) + I(P)) \end{cases} \qquad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} I(R) \\ I(S) \\ I(T) \\ I(P) \\ I(Q) \end{pmatrix} = \frac{1}{K} \begin{pmatrix} I(R) \\ I(S) \\ I(T) \\ I(P) \\ I(Q) \end{pmatrix}$$

Ezagun dugu *M* dela sistemaren matrizea, hau da, grafo zuzenduaren albokotasun matrizea!, eta, baldin eta *I = (I(R); I(S); I(T); I(P); I(Q))'* bada garrantzien bektorea, hau dugula:

$$MI = \frac{1}{K} I$$

alegia, garrantzien bektorea *M* matrizearen bektore propioa dela, eta $\frac{1}{K}$. balio propioa duela.

> «Errekurtsioaren arazoa matrize jakin baten bektore propioen kalkulu bihurtzen da; garrantzien bektorea bektore propioa da».

Aurkikuntza bikain honi eutsiko diogu azterketa osoan.

Ez da komenigarria, ordea, grafo zuzenduaren *M* albokotasun matrizea erabiltzea. Izan ere, ekuazio linealetako aurreko sistema horrek, $K \neq 1$ denean, soluzio bakar bat dauka: soluzio nabaria, hau da, bost orrialdeen garrantzia 0 da; eta baldin eta $K = 1$ bada, soluzioa $I(R) = I(S) = I(T) = 0$, eta $I(P) = I(Q)$ da. Bi aukeretako bat bera ere ez da aski ona.

Gure analisia hobetu behar dugu gure asmoetarako egokiagoa den matrize bat nahi badugu, eta grafo zuzenduari, hots, *M*-ri buruzko informazioa ematea nahi badugu.

2.2. Ausazko nabigatzailea

Orain gogoeta egingo dugu *ausazko nabigatzaile*[14] batek sarean duen portaerari buruz. Demagun gure nabigatzailea P_j orrialdea kontsultatzen ari dela, eta orrialde horrek n_j esteka dituela. Handik aldi batera, P_j orrialdeak duen $I(P_j)$ garrantziaren proportzionala den aldi batera, nabigatzaileak esteka bat aukeratzen du ausaz P_j-tik ateratzeko; esteka horrek P_i orrialdera eramaten du; P_i-ra joateko probabilitatea $\frac{1}{n_j}$ da. Interpretatzen dugu $\frac{I(P_j)}{n_j}$ dela P_j orrialdeak P_i-ri transmititzen dion garrantziaren neurria. Sortzen den matrize berria *M* matrizearen alterazio bat da:

$$
\begin{array}{cc}
 & \begin{array}{ccccc} P_1 & \cdots & P_j & \cdots & P_n \\ \downarrow & & \downarrow & & \downarrow \end{array} \\
\begin{array}{c} P_1 \rightarrow \\ \vdots \\ P_i \rightarrow \\ \vdots \\ P_n \rightarrow \end{array} &
\begin{pmatrix} \vdots & \cdots & \vdots & \cdots & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & \cdots & \frac{b_{ij}}{n_j} & \cdots & \vdots \\ \vdots & & \vdots & & \vdots \\ \vdots & \cdots & \vdots & \cdots & \vdots \end{pmatrix}
\end{array}
\qquad
\begin{array}{l}
b_{ij} = \begin{cases} 1 & \text{baldin eta estekaren bat badago } P_j\text{-}tik\ P_i\text{-}ra \\ 0 & \text{baldin eta estekarik ez badago } P_j\text{-}tik\ P_i\text{-}ra \end{cases} \\
n_j = \text{esteka kopurua } P_j \text{ orrialdean} \\
\quad (n_j \neq 0)
\end{array}
$$

H esaten badiogu matrize berriari, hona hemen zein den ekuazioa definitzeko P_i-ri esleitzen zaion garrantzia:

$$
I(P_i) = K \left[\sum_{P_j\text{-}k \text{ esteka dauka } P_i\text{-}ra} \frac{I(P_j)}{n_j} \right] \qquad (K > 0) \qquad (2)
$$

14. *Random surfer* ingelesezko testuetako adierazpenaren itzulpena da.

Ekuazio linealetako sistemaren matrize-adierazpena (2) $H\ I = \lambda I$ motakoa da, eta adierazten digu garrantzien bektorea H matrizearen bektore propio bat dela.

Estekarik gabeko orrialdeei dagozkien H-ren zutabeak nuluak dira. Izan ere, estekarik gabeko orrialde asko daude *web*-ean. *Web*-aren berezko mintzairan, beste orrialde batzuetara estekarik ez duten orrialde hauek *orrialde esekiak*[15] direla esaten da. Zer egin dezake ausazko nabigatzaile batek estekarik gabeko orrialde batean sartzen denean? Demagun une horretan orrialde bat aukeratzen duela ausaz. Hori eginez gero, orrialde guztietara estekak izango balituzte bezala jokatzen dute estekarik gabeko orrialdeek. Horrela M albokotasun matrizean zutabe nulua zena, koefiziente guztiak $\frac{1}{n}$ (n =*web* orrialdeen kopurua guztira) dituen zutabe bihurtzen da, eta hori probabilitate-banaketa uniformea da.

Behatzen ari garen matrizearen beste alterazio bat eragiten du horrek. $S = (s_{ij})$ deituko diogu matrize berriari:

$$s_{ij} = \begin{cases} \frac{1}{n_j} & \text{baldin eta } P_j \text{ orrialdeak } n_j \neq 0 \text{ esteka baditu, eta esteka bat badago } P_j\text{-tik } P_i\text{-ra,} \\ 0 & \text{baldin eta } P_j \text{ orrialdeak } n_j \neq 0 \text{ esteka baditu, eta estekarik ez badago } P_j\text{-tik } P_i\text{-ra,} \\ \frac{1}{n} & \text{baldin eta } n_j = 0\text{: } P_j \text{ orrialde eseki bat bada,} \end{cases}$$

eta berriz ezartzen badugu I garrantzien bektorea S-ren bektore propioa dela:

$$SI = \lambda I.$$

Askietsiko dugu emaitza hori? Zoritxarrez S matrizeak ez ditu gure azterketarako oinarrizko eskakizunak betetzen. Eskatzen duguna da I garrantzien bektorea S-ren bektore propioa izatea errekurtsioaren arazoa konpontzeko. Ez hori bakarrik; I bektoreak, gainera,

- positibo izan behar ditu osagai guztiak, orrialdeen garrantziak horrelakoak izan behar direla uste baitugu,
- S-ren balio propio positibo batekin lotua egon behar du,
- bektore propio positibo hori bakarra izatea interesatzen zaigu, *web* orrialde bakoitzari esleituriko garrantzietan anbiguotasunik egon ez dadin.

Deusek ez digu esaten S-k bektore propio hori daukanik. Une egokia da Aljebra Lineala agertokian sartzeko.

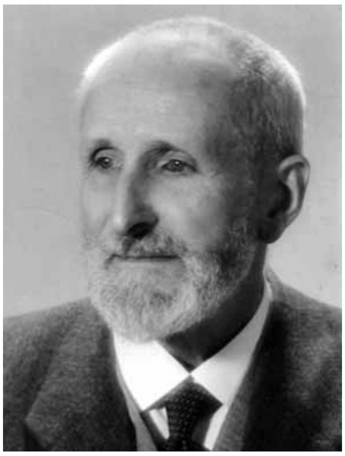
15. Ingelesez estekarik gabeko orrialdeei *dangling nodes* esaten zaie. Grafo zuzenduen teoria orokorreko *putzuei* dagozkie (ingelesez sinks).

2.3. Benetako protagonista: Perron eta Frobeniusen teorema

Bi matematikari alemanek XX. mendearen hasieran formulatutako teorema pare batek ematen digu oinarria algoritmo honek funtzionatzeko: Perron eta Frobeniusen teoremak.

Oskar Perronek[16] 1907. urtean Algebra Linealaren teorema bikainenetako bat demostratu zigun. Teorema horrek erakusten digu nolakoak diren sarrera guztiak positiboak dituen matrize baten bektore propioak eta balio propioak. Hipotesi horren bidez Perronen teoremak erantzuna emango lieke gure eskakizun guztiei. Egia da S-ren sarrera bat bera ere ez dela negatiboa, baina sarrera nulu asko eta asko ditu. Ezin dugu Perronen teorema aplikatu.

Ez da erraza, ezta garrantzirik gabea ere, Perronen teorema zenbaki erreal positibo edo nuluko terminoak dituzten matrizeetan erabiltzea. Hala ere, Ferdinand Georg Frobenius jaunak[17] ekin zion lan horri. Zeregin horretan aritu zen 1908tik 1912era arte, eta teorema bikain hau izan zen lortu zuen emaitza. Ohartarazten dizuet aurkezpen hau oso laburra dela; enuntziatu oso bat eta demostrazio zorrotz bat kontsultatu nahi izanez gero, Aljebra Linealeko hainbat testu daude. Esaterako, C. D. Meyerrena[18]. B. Hupperten[19] testuan azaltzen da H. Wielandten demostrazioa.



Oskar Perron

Ferdinand George Frobenius

16. Oskar Perron matematikari alemana 1880ko maiatzaren 7an jaio zen Frankenthalen, eta 1975eko otsailaren 22an hil zen Munichen.

17. Ferdinand Georg Frobenius Charlottemburgen jaio zen 1849ko urriaren 26an, eta Berlinen hil 1917ko abuztuaren 3an. XX. mendeko matematikari garaienetakoa da, eta oinarrizko ekarpenak egin zizkien Taldeen Teoriari, Algebra Linealari eta Ekuazio Diferentzialen Teoriari.

18. C. D. Meyer: *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, 2000.

19. B. Huppert: *Angewandte Lineare Algebra*, Walter de Gruyter & Co., Berlin, 1990.

Teorema 2.1 (Perron-Frobenius). *Baldin eta A matrize karratu laburtezin bat bada, zeinaren terminoak zenbaki erreal positiboak edo nuluak baitira, honako baieztapen hauek egiaztatzen dira.*

1. *A matrizeak λ balio propioa dauka* Perronen erroa du izena. *Balio horretan*

   a) λ zenbaki erreal positibo bat da,

   b) baldin eta μ A-ren beste balio propio bat bada (agian konplexua), orduan $|\mu| \leq \lambda$,

   c) baldin eta modulu maximoduneko balio propioak baldin badaude, orduan $x^k - \lambda^k = 0$-ren soluzioak dira,

   d) λ A-ren balio sinple bat da: horrek esan nahi du A-ren bektore propio guztiak, λ-rekin lotuak daudenak, oinarrizko bektore bakar baten multiplo eskalarrak direla.

2. *Bada $X = (x_1, ..., x_n)'$ bektore propio bakar bat, A-rena dena, eta balio propioa duena, hots, $AX = \lambda X$, zeinaren koordenatuak denak baitira positiboak, eta halakoa non*

   $\sum_{i=1}^{n} x_i = 1$ *baita.*

   Bektore horri honela esaten zaio: Perronen bektore propioa.

3. *Baldin eta Y bada A-ren beste bektore propio ez negatibo bat, orduan $Y = tX$ da, t eskalar erreal positiboren batentzat.*

Perronen bektorea litzateke gure arazoaren konponbidea. Baina funtsezko bi oztopo ditugu gure bidean. Perron-Frobeniusen teoremak postulatu gisa proposatzen du Perronen bektore propioa dagoela matrize laburtezinetarako. Bestalde, modulu beraren *k* balio propio existitzea ez da desiragarria.

Zer da matrize laburtezin bat? Matrize karratu bat *laburgarria* dela esaten dugu errenkaden eta zutabeen permutazio bat egin eta gero, permutazio bera bi kasuetan, kutxatan deskonposatzen bada.

$$\begin{pmatrix} A_{11} & 0 \\ A_{12} & A_{22} \end{pmatrix}$$

non A_{11}, A_{22} matrize karratuak baitira. Permutaziorik ez badago, *A laburtezina* dela esango dugu.

Gure adibideko *M* matrizeari begiraldi bakar bat egin, eta, horrenbestez, *H* eta *S* matrizeei ere bai, harengandik abiatuz eraiki baitira, eta ikusiko dugu laburgarriak direla. *M*-n agertzen den zeroen kutxa, eta *H*-k eta *S*-k harengandik jasotzen dutena, izatearen arrazoia da ez dagoela *P* edo *Q* erpinetako batean hasten den eta gainerako erpinetako batean amaitzen den bide zuzendurik. Grafoen teoriaren arabera, gure adibideko grafoa ez da oso konexua. Grafo bat konexua da baldin eta bere zeinahi erpinetako bi arkubide baten bidez badaude lotuak, eta grafo zuzendu bat *oso konexua* da baldin eta erpin bakoitza arku-bide baten bidez irisgarria bada beste edozein erpinetatik. Hain zuzen, hurrengo emaitza hau frogatzen da.

Teorema 2.2. *Grafo zuzendu bat oso konexua da baldin eta soilik baldin bere albokotasun matrizea laburtezina bada.*

Horrek esan nahi du Perron-Frobeniusen teorema sarearekin lotutako grafo zuzenduari aplikatzeko, eta, hortaz, dena ongi ibiltzeko, baldintza ezinbestekoa eta nahikoa dela modua edukitzea edozein orrialdetatik beste edozein orrialdetara iristeko, esteken bidez nabigatzen. Baina ezinezkoa da hori. *Web* sarea konexua ere ez da eta!

2.4. Azken aldaketa: Google-ren matrizea

Berriro gogoeta egingo dugu gai honi buruz: nola mugitzen den gure ausazko nabigatzailea sarean. S-k deskribatzen du gure nabigatzailearen portaera: edo dagoen orrialdearen esteka bati jarraitzen dio, edo ausaz aukeratutako beste orrialde bati jarraitzen dio orrialde eseki batean baldin badago. Apur bat aldatuko dugu portaera hori.

α parametro bat aukeratuko dugu 0 eta 1-en artean, eta joko dugu ausazko nabigatzailea, P orrialde bakoitzean, $1 - \alpha$ -ren berdina den probabilitatearekin aspertzen dela, eta, hori gertatzen denean, nabigatzaileak orrialde bat aukeratzen duela ausaz, bai P orrialdearekin lotuta badago, eta ez badago ere, eta nabigatzen jarraitzen duela[20].

Ikusiko dugu ezen, lehen bezala, P orrialde eseki bat bada, nabigatzaileak sareko orrialde guztien artean bat aukeratuko duela ausaz.

Beste hitz batzuk erabiliz, ausazko nabigatzaileak:

- α probabilitatea baldin badu, S matrizea izango du gidari, eta
- $1 - \alpha$ probabilitatea baldin badu, orrialde bat ausaz aukeratuko du.

20. Ingelesezko testu batzuetan parametroari *dumping factor* deitzen zaio.

Google-ren matrizea hau da:

$$\boxed{G = \alpha S + (1 - \alpha)\, E_n}$$

eta bertan E_n da sarrera guztiak $\frac{1}{n}$ -ren berdinak dauzkan matrizea, eta n *web* orrialde kopurua guztira.

$$E_n = \frac{1}{n} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix}$$

Iritsi al dugu gure helburua? Baldin eta $\alpha < 1$ bada, orduan G sarrera guztiak hertsiki positiboak dira; beraz G matrize laburtezin bat da, eta Perron-Frobeniusen teoremaren hipotesi guztiak betetzen ditu (Perronen teoremaren hipotesiak ere betetzen ditu) eta ondorio guztiak egiaztatzen dira ezinbestean. Aurkeztu ditugun emaitza teoriko zoragarriek egiaztatzen dute existitzen dela buru-belarri bilatzen aritu garen garrantzi-bektore hori: Google-ren G matrizeak duen Perronen bektorea da.

Nola kalkulatu, ordea? Ohartaraz dezagun *web* orrialdeen kopurua izugarria dela; gaur egun jotzen da edo, hobeki esateko, susmatzen da «ikusgai» dauden web-orrialdeen kopuruak berrogeita hamar mila milioi inguru egiten duela[21] (hots, $n = 5 \times 10^{10}$). Berebiziko matrize horretan ezinezko zeregina da bektore propioak zuzenean, metodo teorikoaren bidez, aurkitzea. Gure ordenagailuen kalkulu gaitasuna, Borgesek esango zukeen bezala, sekulakoa da, baina finitua.

Zorionez G matrizearen egitura berak kalkulu metodo dotore, eta, batez ere, egingarria ematen digu: G matrizea *estokastikoa* da.

21. Zaila da finkatzea zehazki zenbat *web* orrialde dauden. Helbide honetan: <http://stooge.hubpages.com/hub/How-many-webpages-do-you-think-actually-exist-on-the-Internet> zenbait kalkulu aurkituko dituzue, eta horietatik aukeratu dugu testuan duzuena. Erreferentzia honetan bertan, egileak dokumentuetan oinarritu gabe aieruz esaten du «*internet* ikustezina», hau da, indexatu gabeko orrialdeen kopurua 100 aldiz handiagoa dela, gutxienez.

Argigarria da egiaztatzea, «1» kontsulta egiten badugu Google bilatzailean, nola esango digun emaitzetan 25.270.000.000 inguru erantzun dauzkala.

2.5. Markoven prozesu bat garrantzien bektorea lortzeko

$G = (g_{ij})$ idazten badugu, honako hau izango dugu

$$g_{ij} = \begin{cases} \frac{\alpha}{n_j} + \frac{1-\alpha}{n} & \text{baldin eta esteka bat badugu } P_j\text{-tik } P_i\text{-ra,} \\ \frac{1-\alpha}{n} & \text{baldin eta } P_j \text{ ez bada orrialde esekia, baina ez bada estekarik } P_j\text{-tik } P_i\text{-ra,} \\ \frac{1}{n} & \text{baldin eta } P_j \text{ orrialde esekia bada.} \end{cases}$$

Kalkulu erraz batek erakusten digu zutabe guztietako osagaien batura 1 dela. Beraz, zutabe bakoitza probabilitate banaketa bat dela interpretatzen dugu, halako eran non gure ausazko nabigatzaileak g_{ij} baitu probabilitatea P_j orrialdetik P_i orrialdera lekualdatzeko. Matrize baten zutabe guztiak bektoreak badira, eta bektoreen osagai ez negatiboen batura 1 bada, matrizeari *matrize estokastikoa* esaten zaio. Hau da, G matrizea matrize estokastikoa da.

Nahi izanez gero, arrazoibide hori *Markoven kateak* direlakoen bidez formulatzen dugu. Une bakoitzean ausazko nabigatzaileak.

1. aukeratu egiten du zein orrialdetara lekualdatu, eta horretarako uneko orrialdea bakarrik hartzen du kontuan, eta ez bisitatu dituen aurreko orrialdeak, hau da, memoriarik gabe aukeratzen du, eta
2. P_j orrialdetik P_i orrialdera igarotzeko probabilitatea konstantea da beti.


A. A. Markov

Horra hor, hain zuzen, zein ezaugarrik definitzen dituzten *Markoven kateak*[22].

Hau da, G matrizea matrize estokastikoa da, eta gure Markoven katearen *trantsizio matrizea* da. Interpretazio horren arabera, $(i; j)$ tokia hartzen duen eskalarra G^m potentziadun matrizean, honela interpretatzen da: ausazko nabigatzaileak duen probabilitatea igarotzeko P_j orrialdetik P_i orrialdera, m urratsean. Horrenbestez, interpretazio honen arabera, orrialde garrantzitsuenak dira ausazko nabigatzaile batek bisitatuak izateko probabilitate handiena dutenak.

Sarrera guztiak positibo dituen matrize estokastiko bati Perron-Frobeniusen teorema aplikatzean, propietate adierazgarri bat aurkituko dugu.

22. *Andrei Andreyevich Markov* errusiar matematikaria Ryazan-en, Errusia, jaio zen 1856ko ekainaren 14an eta Petrogradon (Orain San Petersburg), Errusia, hil zen 1922ko uztailaren 20an.

Teorema 2.3. *Baldin eta G bada sarrera guztiak positibo dituen matrize estokastiko bat, G-k duen Perronen erroa 1 da.*

Gainera, Perronen erroa 1 moduludun G-ren balio propio bakarra da.

Potentzien metodoa, edo Markoven metodoa izenekoa, aplikatzeko, eragiketa hauek egingo ditugu:

1. I_0 probabilitate-bektore bat aukeratuko dugu izangai I izateko.

2. Hurrenez hurren kalkulatuko ditugu:

$$
\begin{aligned}
&I_1 = GI_0,\\
&I_2 = GI_1 = G^2\, I_0,\\
&I_3 = GI_2 = G^3\, I_0,\\
&\ldots
\end{aligned}
$$

eta, oro har, $I_{k+1} = GI_k = G^{k+1} I_0$.

Potentzien metodoa eraginkorra izateko eta nahi dugun emaitza lortzeko, segida horrek konbergentea izan behar du G-k duen Perronen bektorearekin, hori baita bilatzen duguna, alde batera utziz zein den prozesua hasteko hartu dugun I_0 bektorea. Hurrengo teoremak erantzun osoa emango digu

Teorema 2.4. *Baldin eta G bada sarrera guztiak positibo dituen matrize estokastiko bat, orduan*:

1. *konbergentea da $\{G^k\}$ segida, eta baldin eta $\lim_{k\to\infty} G^k = L$ bada orduan L matrize estokastikoa da;*

2. *L-ren zutabe bakoitza G-k duen IPerronen bektorea da;*

3. *zeinahi I_0 probabilitate-bektoretarako, hau egiaztatzen da:*

$$\lim_{k\to\infty} (G^k I_0) = I.$$

Egin duguna uler dezagun, geneukan helburua erdietsi dugu: I bektorea, G-k duen Perronen bektorea, Markoven potentzien metodoa aplikatzean agertzen diren bektoreen muga gisa agertzen dena, hasieratik bilatzen aritu garen garrantzien bektorea da.

Merezi du nabarmentzea zein den gakoa zeinari esker frogatzen baitugu konbergentea dela G^m potentzien segida: G^m potentziadun matrizearen balioak G-ren balio propioen potentziak dira. Perronen balioa salbuetsita, 1 baita, gainerako balio propio guztiek hertsiki 1 baino txikiagoa den modulua dute, beren hurrenez hurrengo potentziak gero eta txikiagoak dira, hau da, zerorantz doaz. Horren ondorioz, gainera, G-ren bigarren balioa zenbat eta txikiagoa izan, hainbat lasterrago gertatzen da konbergentzia[23]. Egoki da hemen hau esatea: S matrize estokastikoaren balio propioak baldin badira $\{1, \lambda_2, ..., \lambda_n\}$, orduan $G = \alpha S + (1 - \alpha)E_n$ matrize estokastikoaren balio propioak dira $\{1, \alpha\lambda_2, ..., \alpha\lambda_n\}$[24]. Izan ere, k iteraziotan egiten den akatsa $2\alpha^k$ baino handiagoa da. Hortaz, zenbat eta txikiagoa izan, orduan eta lasterragoa izango da konbergentzia.

Baina, bestalde, α 0-ra hurbiltzen bada, interpretazio hau egiten dugu: ausazko nabigatzailea aspertu egin da probabilitate handiago batekin eta sarearen egituraz ahantzi da. Izan ere, baldin eta $\alpha = 0$ bada, orduan G nabigatzaile baten matrize bat da, zeinetan orrialde bakoitzak ausaz aukeratzen baitu hurrengo orrialdea, esteka egon edo ez egon.

Esperimentu asko egin eta gero, Google-k hau hartu zuen:

$$\boxed{\alpha = 0,85}$$

Balio horrekin, 50 eta 100 iterazio bitarteko zenbaki batean I-ren hurbilketa on bat lortzen dugu.

Zehaztu dugu, beraz, Google-ren *matrizea* esaten diotena:

$$\boxed{G = 0,85S + 0,15E_n}$$

23. Bigarren balio propioaren garrantziaren gaia azaltzeko, eta ez horretarako soilik, oso da gomendagarria K. Bryant eta T. Leiseren lan hau: *The $25,000,000,000 eigenvector, The Linear Algebra behind Google*. SIAM Rev., 48(3) (2006) 569-581.

Kontsultatzea merezi du, halaber, T. H. Haveliwala eta S. D. Kamvar-en lan interesgarri hau: *The Second Eigenvalue of the Google Matrix*. Technical Report, Stanford, 2003. <http://ilpubs.stanford.edu:8090/582/1/2003-20.pdf>.

24. Badago egitate horren demostrazio bat A. Langville eta C. Meyerren honako liburu bikain eta atsegin honen Theorem 4.7 izenekoan: *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006.

Hain zuzen ere, G-k duen Perronen bektorea *PageRank*-en garrantzien bektorea da. Kalkulu eraginkorra egiteko, Markoven kate bat behar dugu, zeinaren trantsizio-matrizea G izango baita[25].

Azkenik, halako tamaina erraldoidun matrize batekin egiten den konputazio eraginkorrari buruzko zalantza batzuk argitu behar ditugu metodo honen bidez. S matrizea, arrazonamendu honen tarteko fasean agertu zena, batuketa bat da: $S = H + A$, zeinean A matrize bat baita, halakoa non:

- H-ren zutabe ez nuluak zutabe nuluak baitira A-n;
- H-ren zutabe nuluek dauzkaten koefiziente guztiak baitira $\frac{1}{n}$ A-n.

Hau daukagu, beraz,

$$GI_k = 0{,}85HI_k + 0{,}85AI_k + 0{,}15E_nI_k$$

- H-ren sarrera gehienak nuluak dira. Hain zuzen, *web* orrialdeen batez besteko esteka kopurua 10 da; hau da, zutabe bakoitzak hamar bat sarrera ez nulu ditu bakarrik. Hortaz, HI_k kalkulua egiteko hamar bat kalkulu egin behar dira bakarrik.

25. Bitxikeria modura, esan dezagun Google-ren matrizea, adibide gisa hartutako sare nimiñoari lotua, hau dela:

$$G = \begin{pmatrix} 3/100 & 1/5 & 3/100 & 3/100 & 3/100 \\ 97/400 & 1/5 & 91/200 & 3/100 & 3/100 \\ 97/400 & 1/5 & 3/100 & 3/100 & 3/100 \\ 97/400 & 1/5 & 3/100 & 3/100 & 22/25 \\ 97/400 & 1/5 & 91/200 & 22/25 & 3/100 \end{pmatrix}$$

Matrize hau oso txikia denez, Perronen bektorea zuzenean kalkulatuko dugu MATHEMATICA.8 erabiliz. Hona hemen emaitza:

$$I' = (I(R), I(S), I(T), I(P), I(Q))' = \left(\frac{9600}{226007}, \frac{16587}{226007}, \frac{11640}{226007}, \frac{3431860}{8362259}, \frac{3530800}{8362259}\right)'$$

eta bost orrialdeen ordenazio hau dakar:

$$I(Q) > I(P) > I(S) > I(T) > I(R) \Rightarrow Q \succ P \succ S \succ T \succ R.$$

G trantsizio matrizea daukan Markoven prozesu baten hirurogeita zortzi iterazio egin eta gero, Perronen bektorearen hurbilketa hau aurkitu dugu:

$$I' \approx (0{,}0424766,\ 0{,}0733915,\ 0{,}0515028,\ 0{,}410399,\ 0{,}42223)'$$

Balio hurbilduak hartzean egindako akatsa formula honek bornatzen du:

$$2 \times 0,\ 8568 \approx 3{,}17 \times 10^{-5}.$$

Hain zuzen, zuzenean kalkula daiteke errorea $4{,}1 \times 10^{-7}$ baino txikiagoa dela.

- AI_k kalkuluak ematen digu batugaia «orrialde esekien» garrantziarekin. A-ren errenkada guztiak berdinak direnez, batugai hori balioztatzeko kalkulu bakarra egin behar da.

- Antzeko zerbait gertatzen da E_nI_k-rekin; bere kalkuluak batugaia ematen digu *web* orrialde guztien garrantziarekin batera. Aurreko kasuan bezala, E_n-ren errenkada guztiak berdinak direnez, batugai hori balioztatzeko kalkulu bakarra egin behar da.

3. Epilogoa

0,85 balioa parametrorako ezartzea erabaki enpirikoa da soil-soilik. Ez da bakarra: E_n matrizearen aukeraketa ere hala da: 1 heina daukan beste matrize estokastiko bat, probabilitate banaketa ez homogeneo bat diren zutabeak izango zituena, erabil zitekeen. Aitzitik, *PageRank*-en postulatu nagusiari datxikion errekurtsioaren arazoa, bektore propioen bidez, saihesteko modua ez da enpirikoa. Ezta Markoven kateen arrakasta ere, Perron- Frobeniusen teoremen erabileran oinarritua, garrantzien bektorea zehazteko. Algoritmo harrigarri eta etekintsu honen sortzaileek erabilitako matematikaren ezagutza sakona zuten, eta, hortaz, zailtasunak gainditu eta beren arrazoibidea hobetu ahal izan zuten, eraikuntza eraginkorra lortzeko. Ez zuten aski izan formula pare bat erabiltzea, baizik eta, nola sortu zen eta zertan oinarritzen zen ulertuz, gai izan ziren eredu erabilgarri bat sortzeko, hipotesi teorikoak zorrotz beteko zituena, emaitza bermatzeko.

Batzuetan, Matematikari eta bereziki Aljebrari bigarren mailako zereginak ematen zaizkie teknologien munduan. Batzuetan harrokeria akademikoz puztutako baieztapenak entzuten dira, honen gisakoak aldarrikatzeko: «azalduko dugu guk matematika, guk baitakigu zer behar den», eta jarraian, «funtsezkoa da irakasgaiek alderdi praktiko bat izatea». Ez dut nik ukatuko praxiaren garrantzia, baina ez du ezertarako balio zerbait nola egiten den jakiteak, zer den ez badakigu.

Titulazio bateko egiazko praktikek ibilbide profesional osoa iraungo dute. Ikaskuntza urte hauetan ez badiegu gure ikasleei oinarri teoriko sendo bat ematen, beren bilakaera profesionalean izango dituzten erronka anitzetan aplikagarria izango dena, litekeena da gure tituludunak gidaliburuak irakurtzeko gai izatea, baina ez dira gidaliburuak idazteko gai izango. Ezagutza hutsari uko egitea gure etorkizun intelektuala mengeltasunera eramatea da.

Esan beharra esan dut.