NYILATKOZAT

Név: Kelemen Lajos

ELTE Természettudományi Kar, szak: Matematika BSc

NEPTUN azonosító: G6WPAB

Szakdolgozat címe:
A bitcoin tranzakciók eloszlása és skálázódása

A **szakdolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2023. 06. 03.

Kelemen Lajos

a hallgató aláírása

Eötvös Loránd Tudományegyetem
Természettudományi Kar

Kelemen Lajos
Matematika BSc
alkalmazott matematikus szakirány

A bitcoin tranzakciók eloszlása és skálázódása

Szakdolgozat

Belső konzulens: Dr. Backhausz Ágnes egyetemi adjunktus
Valószínűségelméleti és Statisztika Tanszék

Külső témavezető: Seres István András doktorandusz
ELTE IK Komputeralgebra Tanszék

Budapest, 2023.

Tartalomjegyzék

Köszönetnyilvánítás

Köszönettel tartozom külső témavezetőmnek, Seres Istvánnak, hogy megismertette velem ezt a témát, és az ötlete alapján elkészíthettem a szakdolgozatomat. Szaktudásával rengeteg segítséget nyújtott a kutatás során, amely által sok új ismerettel lettem gazdagabb.

Hálás köszönettel tartozom továbbá belső konzulensemnek, Backhausz Ágnesnek, hogy ötletei jóvoltából más megközelítést nyújtott a téma felé, folyamatosan segítette a szakdolgozatom létrejöttét, és hogy mindig a rendelkezésemre állt.

Szeretném kifejezni hálámat a családom részére, akik mindvégig támogattak és támaszt nyújtottak a tanulmányaim során.

Végül, de nem utolsósorban köszönetemet fejezem ki kollégiumi és egyetemi barátaim felé, akikkel közösen tölthettem az elmúlt éveket, és bármilyen kérdéssel kapcsolatban a segítségemre siettek.

Bevezetés

A szakdolgozatom a bitcoin tranzakciók különböző tulajdonságainak vizsgálatát írja le a rendelkezésemre bocsájtott tranzakciós adatok alapján. A hálózat feltérképezéséhez elengedhetetlen a Bitcoin rendszerének alapvető ismerete. Az első fejezetben ismerkedhetünk meg ezekkel a fundamentumokkal. Fontos megérteni, hogy a Bitcoin újdonsága, hogy egy decentralizált pénzügyi rendszert hoz létre, amelynél nincs szükség felügyelő hatóságra. A rendszer működésének leírásából kiderül, hogyan jönnek létre a tranzakciók, és hogy mi is valójában egy bitcoin, valamint az is, hogyan épül fel egy blokk. A fejezet végén betekintést nyerünk a Bitcoin eddigi felhasználásának módjaiba, és megismerjük, hogy mik a rendszer mozgatórugójai.

A második fejezetben a tranzakciós gráfon keresztül érthetjük meg a hálózat működését. Gráfelméleti vizsgálatok segítségével próbáljuk egy jól ismert hálózattal párhuzamba állítani a Bitcoint. Fontos tulajdonságokat sikerül belátni a gráffal kapcsolatban, ilyen például a skálafüggetlenség bizonyítása. A gráf különböző élsúlyainak eloszlásából láthatjuk, hogy milyen fontos szerepet játszanak a különböző szolgáltatók. A tranzakciók értékeit, idejét és a bitcoinok árfolyamát tartalmazó ábrák elemzése pedig tovább segíti a megértést.

A harmadik fejezetben ismerkedhetünk meg a kutatás movitációját adó cikkel, ami a készpénz mozgásának leírásával foglalkozik. A cikk szerint a bankjegyek sokáig egy kis sugarú környezetükben mozognak, majd egy nagyobb ugrás keretében kerülnek messzebb, nem pedig fokozatosan távolodnak. A bitcoinok esetében más eloszlás figyelhető meg, sokkal jelentősebbek a nagy távolsággal rendelkező tranzakciók. A távolság és a tranzakció többi paraméterének korrelációja közül az értéknél kapott eredmény szerint korrelálatlanok, tehát nincs közöttük összefüggés. Szórásanalízis alkalmazásával információt kapunk arról, hogy a kontinenseken belüli tranzakciók távolságában és értékében van-e szignifikáns eltérés. A fejezet legvégén a sok tranzakcióval rendelkező Bitcoin címek deanonimizálására kerül sor, amivel még jobb képet kapunk a rendszer résztvevőiről.

1. fejezet

A Bitcoin

1.1. Kezdetek

A Bitcoint 2009. január 3-án, a 2008-as gazdasági világválság hatására hozta létre egy ismeretlen személy, akit Satoshi Nakamoto álnéven ismerünk [10]. Ez az első és legismertebb kriptovaluta, valamint a legnagyobb piaci kapitalizációval rendelkező is.

A Bitcoin legfőbb újdonsága, hogy a blokklánc technológia segítségével egy decentralizált rendszert hoz létre, ahol a felhasználók úgynevezett Peer-to-peer módon, azaz közvetlenül egymás között, harmadik fél résztvétele nélkül tudnak tranzakciókat lebonyolítani. Így tehát nincs szükség egy felügyelő hatóságra, aki ellenőrzi a folyamatot. A blokkláncra kerülő tranzakciók ugyan publikusak, így a résztvevők címei is ismertek, de a valós kilétüket nem tudjuk beazonosítani. Ezt pszeudoanonimitásnak nevezzük. Csak ezen adatok ismeretében tehát nem biztos, hogy azonosítani tudjuk a cím valós tulajdonosát, de ha egyszer megtettük, akkor onnantól fogva az összes korábbi tranzakciója is ismertté válik.

A rendszer működésének leírása Satoshi Nakamoto cikke [10] és a *Bitcoin and Cryptocurrency Technologies* című könyv [11] alapján készült. A nagy kezdőbetűs Bitcoin alatt általánosságban a rendszert értjük, míg a kis kezdőbetűs bitcoin a rendszer által használt eszközt jelöli.

1.2. Működése

A blokklánc egy megosztott, nyilvános főkönyv, amelyre a teljes Bitcoin hálózat támaszkodik. Az összes visszaigazolt tranzakció bekerül egy blokkba, amely hozzáadódik a blokklánchoz. Így a Bitcoin walletek (1.7. Definíció) ki tudják számítani elkölthető egyenlegüket, valamint az új tranzakciók esetében visszaigazolható, hogy ténylegesen a tulajdonos által birtokolt bitcoinok elköltése történt-e meg.

A tranzakciók validálását a bányászok végzik a Proof-of-Work konszenzus alapján.

Ennek lényege, hogy a bányászoknak egy nagy számítási kapacitást igénylő feladatot kell megoldaniuk, melyet ha elsőként teljesítenek jutalomban részesülnek, így kerülnek új bitcoinok a rendszerbe. A jutalom eleinte 50 bitcoin volt, viszont körülbelül négyévente feleződik ez a szám, jelenleg 6,25 BTC. Ez a folyamat 2140-ig fog tartani, ekkorra a bitcoinok száma eléri a maximális 21 milliót, ezután a bányászok bevétele a tranzakciós díjakból származik majd.

1.2.1. Hash puzzle

1.1. Definíció. Hash függvénynek nevezünk egy függvényt [11], ha egy tetszőleges hosszúságú karaktersorozathoz, vagyis stringhez, hozzárendel egy fix hosszúságú stringet, és ezt egy n hosszú bemenetre $\mathcal{O}(n)$ időben teszi. Kriptográfiai hash függvény esetén további három kritériumot követelünk meg. Nehéz legyen két olyan stringet találni, amelyeknek megegyezik a hashe, és olyan stringet, aminek egy adott érték a hashe. Emellett ha egy kis változtatást hajtunk végre az inputon, akkor egy teljesen különböző outputot kell kapjunk. Az egyik legismertebb az SHA–256 hash függvény, amelyet a Bitcoin is használ.

A megoldandó feladat, amit hash puzzle-nek is hívnak, egy megfelelő szám, úgynevezett nonce megtalálása. Egy szám akkor lehet nonce, ha a szám, az előző tranzakció hash kódja ($\text{hash}_{\text{previous}}$) és az új blokkba kerülő tranzakciók (tx) listájának konkatenálásával kapott stringre alkalmazva a hash függvényt, a kapott érték egy meghatározott célérték (target) szint alatt van.

$$H(\text{nonce}||\text{hash}_{\text{previous}}||\text{tx}||\text{tx}\ldots) < \text{target} \tag{1.1}$$

Ezt próbálgatással tudják megkeresni a bányászok. Ha valaki megtalálja, akkor onnan már bárki könnyen le tudja ellenőrizni a helyességét, hiszen csak alkalmazni kell rá a hash függvényt. Ilyenkor a bányász a blokk egészét elküldi az úgynevezett node-oknak. A node egy számítógép, amely csatlakozik a Bitcoin hálózatra és tárolja a blokklánc másolatát, valamint validálja a blokkokat, de bányász tevékenységet nem végez, így nem kell nagy számítási kapacitással rendelkeznie. A node-ok validációja után csatlakozik az új blokk a blokkláncra.

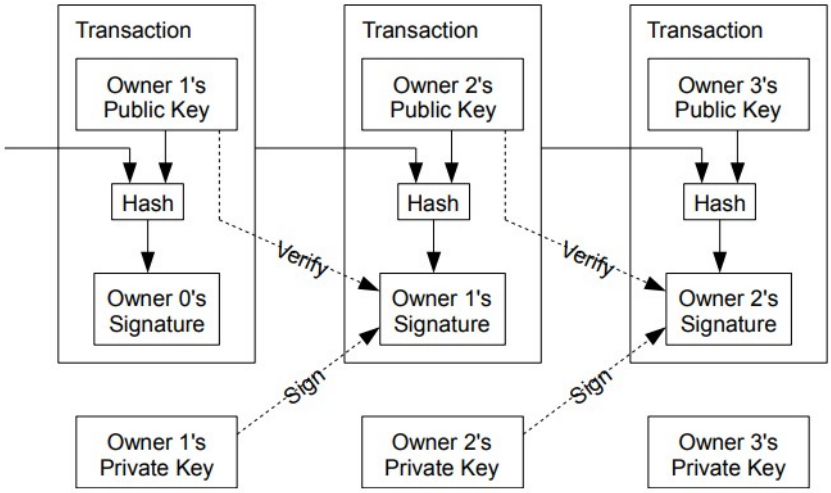
A hash puzzlenek három fontos tulajdonsága van [11]. Elég nehéznek kell lennie, hogy ne kerülhessenek be túl gyorsan a blokkok. Másrészt viszont 2016 blokkonként a rendszer újraszámolja a target értéket úgy, hogy a blokkok megközelítőleg 10 percenként jöjjenek létre. Ez a gyakorlatban megközelítőleg két hétnek felel meg, mivel $2016 \cdot 10$ perc pontosan ennyi idő. A harmadik fontos tényező pedig az ellenőrzés trivialitása, ami a decentralizáltság miatt fontos, mert így bárki képes validálni, nem kell egy központi szervezet.

1.2.2. Bitcoin tranzakció

1.2. Definíció. A digitális aláírás [11] egy olyan metódus, melynek segítségével igazolni tudjuk egy üzenet eredetét, és hogy az aláírás óta nem változott. A metódus során először létrehozunk egy privát és egy publikus kulcsot. Ezután az üzenet és a privát kulcs segítségével létrehozzuk a digitális aláírást. Ennek ellenőrzéséhez pedig szükség van egy hitelesítő függvényre, amely az üzenet, a publikus kulcs és a digitális aláírás felhasználásával dönt a hitelességről. Fontos követelmény, hogy az aláírásunkat ne lehessen hamisítani, tehát egy digitális aláírásunk és a publikus kulcsunk ismeretében ne lehessen más üzenetre létrehozni a digitális aláírást úgy, mintha mi írtuk volna alá.

A bitcoin küldéséhez is szükség van egy privát kulcsra, amit a tranzakciók aláírásához használunk, és egy publikus kulcsra, amit Bitcoin címnek is nevezünk és az aláírások ellenőrzésére használunk. A Bitcoin cím egy véletlenszerűen generált betű– és számsorozat. A privát kulcs egy másik betű– és számsorozat, azonban a Bitcoin címmel ellentétben ez titkos. A Bitcoin címre gondolhatunk egy transzparens széfként, mivel bárki láthatja a tartalmát, de csak a privát kulcs ismeretében juthat hozzá.

Egy bitcoin tulajdonképpen a digitális aláírások egy lánca. A folyamat úgy zajlik, hogy a küldő a coinhoz tartozó előző tranzakció hashét és a kedvezményezett publikus kulcsát digitálisan aláírja a privát kulcsával, és ezt fűzi fel a láncra. Ez lesz elküldve a bányászoknak, akik ezt könnyen ellenőrizni tudják, hogy hiteles-e, mert ezen információk mindegyike publikus, így kerül be a tranzakció a blokkba. Felmerülhet még a double-spending, azaz egyazon bitcoin kétszeres elköltésének a problémája. Ezt egy időbélyeg hozzáadásával küszöböli ki a rendszer.



1.1. ábra. Digitális aláírás a Bitcoin tranzakciónál [10]

Egy tranzakciónak nem csak egy inputja lehet. Ez azt jelenti, hogy egyszerre több címről is lehetséges az utalás, viszont ilyenkor az összes cím privát kulcsának az ismerete szükséges. Ez alapján azt feltételezhetjük, hogy ezen különböző címeknek egy tulajdosona van, mivel a privát kulcs megosztásával veszélybe kerülnek a címhez tartozó bitcoinok.

1.3. Definíció. UTXO - Unspent transaction output: Egy tranzakció során az elköltetlen bitcoinok. Egy UTXO addig létezik, amíg inputként nem szerepel.

1.4. Definíció. Coinbase tranzakciónak nevezzük azokat a tranzakciókat, melyeken keresztül a bányászok megkapják a blokkjutalmukat. Mivel ilyenkor új bitcoinok kerülnek a rendszerbe, ezért nincs inputja, de az output több cím is lehet.

1.5. Következmény. *Minden bitcoin első tranzakciója egy coinbase tranzakció, mivel a bitcoinok csak így kerülhetnek be a rendszerbe.*

1.6. Definíció. Satoshinak nevezzük a bitcoin legkisebb tört egységét, egy satoshi százmilliomod bitcoinnak felel meg. Egy tranzakciónál ez a legkisebb mennyiség, amit küldhetünk.

A Bitcoin tranzakciók nem rendelkeznek azzal a tulajdonsággal, hogy pontosan a kívánt mennyiséget küldik el, ehelyett egy a küldőhőz tartozó, úgynevezett change címre kerülnek vissza a UTXO-k. Így az outputban a kedvezményezett címek mellett megjelenik a change cím is. Természetesen ha a címünkön lévő összes bitcoint szeretnénk elküldeni, akkor nem jön létre change cím. A gyakorlatban ezek a tranzakciók az úgynevezett walletek között jönnek létre, amiket különböző szolgáltatóknál érhetünk el.

1.7. Definíció. A Bitcoin wallet egy olyan fizikai eszköz vagy szolgáltatás, melynek segítségével biztonságban tudható és egyszerűen kezelhető a birtokolt bitcoin. Két fő funkciója közül az egyik, hogy tárolja a felhasználó privát és publikus kulcsát, másrészt pedig a tranzakciók megkönnyítésére szolgál. A walletek természetesen digitálisan és fizikálisan sem képesek a bitcoinok tárolására, mivel azok a blokkláncon vannak.

Megkülönböztetünk hot és cold walleteket. A hot walletek szoftver alapúak és csatlakoztatva vannak a hálózatra, így bár könnyebb velük tranzakciót létesíteni, de kisebb biztonságot nyújtanak. Ezzel szemben a cold walletek egy hardveren tárolják a publikus és privát kulcsokat, ami a legbiztonságosabb lehetőség. Ha viszont tranzaktálni szeretnénk velük, akkor újra csatlakozni kell egy hot wallethez.

1.2.3. Egy blokk felépítése

A blokkok tartalmazzák a tranzakciókat, körülbelül 2700 kerül be egybe. A blokkok a blokkláncra egymás után kerülnek be a megfelelő nonce érték megtalálása és a validáció után. Ha egy támadó egy már felfűzött blokkban szeretne megváltoztatni egy tranzakciót, akkor megváltozik a blokk hashe az (1.1) összefüggés miatt, valamint a következő blokk

hashe is, és ezáltal az összes ezután lévőé is, így ezek mind érvénytelenné válnak. Ez a funkció biztosítja, hogy miután egy blokkot hozzáadtak a blokklánchoz, sem az, sem a benne lévő tranzakciók nem módosíthatók. Egy blokk mérete legfeljebb 1MB lehet, jelenleg a teljes blokklánc nagysága 478GB.

A blokk struktúrája az alábbi módon épül fel [2]:

- A blokk mérete: 4 byte
- Blokk Header: 80 byte
- Tranzakció számláló: 1-9 byte
- A blokkban lévő tranzakciók

Érdemes még megvizsgálni a Header felépítését is:

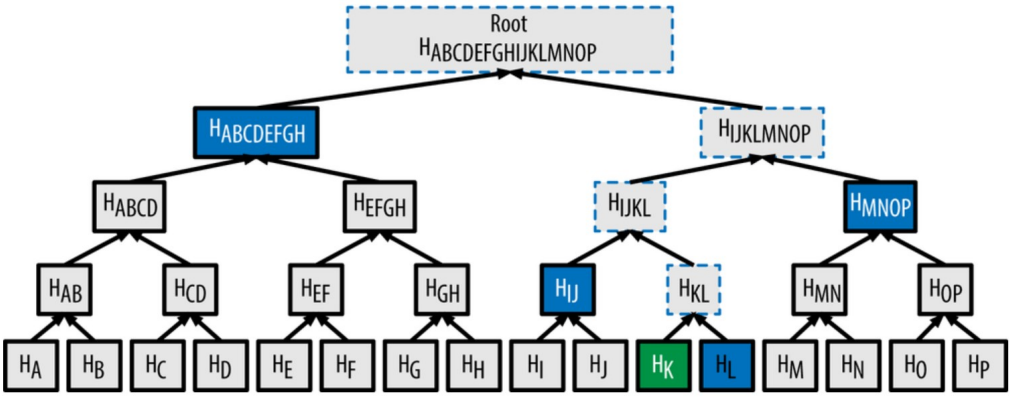
- Verziószám: 4 byte
- Az előző blokk hashe: 32 byte
- Merkle-fa gyökerének hashe: 32 byte
- Időbélyeg: 32 byte
- Target vagy difficulty értéke: 4 byte
- Nonce: 4 byte


```
"size" : 43560,
"version" : 2,
"previousblockhash" :
    "0000000000000027e7ba6fe7bad39faf3b5a83daed765f05f7d1b71a1632249",
"merkleroot" :
    "5e049f4030e0ab2debb92378f53c0a6e09548aea083f3ab25e1d94ea1155e29d",
"time" : 1388185038,
"difficulty" : 1180923195.25802612,
"nonce" : 4215469401,
"tx" : [
    "257e7497fb8bc68421eb2c7b699dbab234831600e7352f0d9e6522c7cf3f6c77",

[... many more transactions omitted ...]

    "05cfd38f6ae6aa83674cc99e4d75a1458c165b7ab84725eda41d018a09176634"
```


1.2. ábra. Egy blokk felépítése [2]

1.8. Definíció. A Merkle-fa [2], más néven bináris hash fa, egy olyan adatstruktúra, amely nagy adathalmazok hatékony összegzésére és ellenőrzésére használható, ami a Bitcoin esetén a tranzakciókra van alkalmazva. A csúcsokban hash értékek szerepelnek. A levelekben egy-egy tranzakció hashe van, majd két ilyen hash konkatenáltjára alkalmazva a hash függvényt kapjuk meg a szülőhöz tartozó hash értéket. Ezt addig folytatjuk, míg egy értéket nem kapunk. Ez lesz a Merkle-fa gyökere, ami bekerül a blokk Headerjébe. A fa mélysége n tranzakció esetén $\lceil \log_2(n) \rceil$.

A Merkle-fa segítségével könnyen ellenőrizni tudjuk, hogy egy tranzakció valóban szerepel-e az adott blokkban. Ehhez szükség van a tranzakció Merkle útjára, ami a leveltől a gyökérig vezető úton lévő csúcsok másik gyermekét tartalmazza. A példa esetén a H_K-hoz tartozó Merkle út a $H_L, H_{IJ}, H_{MNOP}, H_{ABCDEFGH}$ csúcsok.



1.3. ábra. Merkle-fa és a H_K tranzakcióhoz tartozó Merkle út [2]

Így annak bizonyítására, hogy H_K valóban szerepel a fában mindössze $\lceil \log_2(n) \rceil$-szer, jelen esetben 4-szer kell alkalmazni a hash függvényt. Ha az utolsó hash megegyezik a gyökér hashével, akkor a tranzakció valóban szerepel a Merkle-fában.

1.2.4. A rendszer működésének összegzése

A Bitcoin Proof-of-Work konszenzusa tehát az alábbi módon működik [10]:

1. Az új tranzakciókat megosztjuk a bányászokkal.
2. Mindegyik bányász egy blokkba gyűjti a tranzakciókat.
3. A bányászok keresik a megfelelő nonce értéket.
4. Ha egy bányász megtalálta, akkor megosztja a node-okkal.
5. A node-ok elfogadják a blokkot, ha minden tranzakció valid benne.
6. A blokk csatlakozik a blokklánchoz, és a bányász megkapja a blokkjutalmát.

1.3. Használata

A Bitcoin, mint fizetési rendszer leginkább a készpénzes fizetéshez hasonlítható, tulajdonképpen egy digitális változata. A résztvevőket csak a pszeudoanonimitás erejéig ismerjük és a tranzakciók visszafordíthatatlanok, tehát ha egy felhasználó rossz címre utal, vagy csalás áldozata lesz, akkor nem számíthat kártérítésre. A készpénzzel ellentétben viszont szükség van a tranzakció validálására.

Sok kritika éri a Bitcoint a tekintetben, hogy valóban pénznek számít-e, ugyanis nem teljesíti a pénz közgazdasági definíciójában foglaltakat.

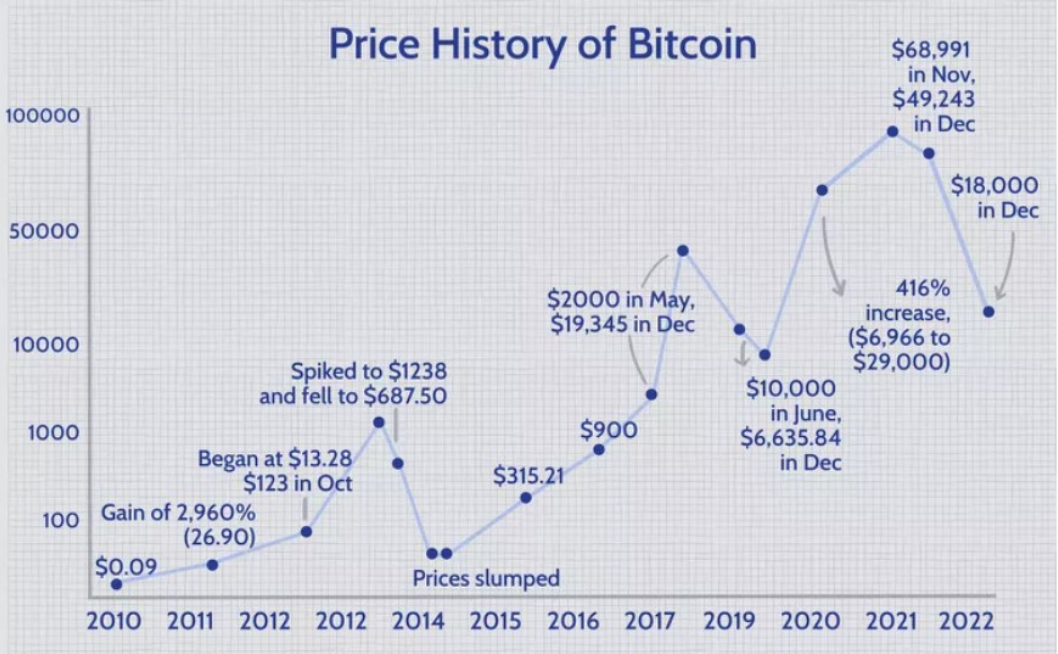
1.3.1. A hálózat résztvevői

A bányászok hatékonyságuk növelése érdekében nem egyedül, hanem úgynevezett Mining pool-okba szerveződnek és a blokkjutalmat elosztják egymás között. Itt hatalmas számítási kapacitás koncentrálódik, ezáltal az energiaigénye is hatalmas méreteket ölt. Egyes országokban, például Kínában, ezért be is tiltották a bányászást.

Fontos szerepet játszanak a kriptotőzsdék, ahol a felhasználók devizára válthatják bitcoinjukat, valamint tőzsdei ügyleteket bonyolíthatnak le. Az egyik legismertebb és legnagyobb ilyen oldal a Binance. A wallet szolgáltatóknál pedig biztonságban tudható, és egyszerűen kezelhető a birtokolt bitcoin. Jelentős részt képeznek a különböző kereskedő oldalak, ahol konkrét termékeket árusítanak. Megemlítendő még a szerencsejátékhoz köthető szolgáltatások, melyek közül a legismertebb a Satoshi Dice [9]. A kereskedésben sokan alkalmazzák a *HODL* néven elhíresült stratégiát, ami a bitcoin jövőbeli magas árfolyamára spekuláló hosszútávú befektetést jelenti.

1.3.2. Változó tendenciák

A 2009-es indulástól kezdődően egészen 2011-ig a tranzakciók jelentős részét a coinbase tranzakciók tették ki, ekkor a blokkjutalom 50 BTC volt. Kereskedés hiányában az árfolyam sem növekedett, 2011 februárjában ért először egy dollárt egy bitcoin. Ezután jelent meg egyre több szolgáltató, amelyek elég kezdetlegesek voltak, 2023-ra kevés kivétellel az összes megszűnt. Nagy port kavart a Mt. Gox tőzsde csődje, ugyanis 2014 elején a tranzakciók közel 70%-át bonyolította le. Ez nagy bizalmatlanságot eredményezett, a bitcoin értéke 1000$-ról 100$ alá csökkent. 2016-ban a média megjelenések hatására (hype) már a 20000$-t is meghaladta az árfolyam, több nagy cég is Bitcoin érdekeltségeket szerzett. 2021-ben El Salvador lett az első ország ahol törvényes fizetőeszközzé nyilvánították a Bitcoint. A legmagasabb értéket 2021 decemberében érte el, ami 68991$ volt. A jövőbeli jelentőségéről megoszlanak a vélemények.



1.4. ábra. A Bitcoin árfolyama 2009 és 2022 között [6]

1.4. A hálózat feltérképezése

A blokkláncról nyert információk segítségével többféle lehetőség nyílik a vizsgálatra. Az egyik megközelítés az ismert címek klaszterezésén alapszik [9]. Ekkor a különböző szolgáltatók publikus címeinek ismeretében heurisztikák alkalmazásával lehet következtetéseket levonni a Bitcoin tranzakciók eloszlásáról, mintázatáról. A tranzakciós gráf felépítésével pedig a hálózat tulajdonságait lehet elemezni. A másik megközelítés a Bitcoin címek deanonimizálása, földrajzi helyzetének meghatározása [7]. Egy valószínűségszámítási modell alapján lehet beazonosítani a felhasználók IP-címét és ennek segítségével a helyzetüket. Ezen adatok birtokában nyomon lehet követni a bitcoinok mozgását térben és időben, valamint a hálózat kiterjedését az egész bolygóra nézve.

Szakdolgozatomban a kapott adatok segítségével, amelyek legtöbbje 2013-as tranzakció, valósítom meg a fent említett vizsgálatokat. A felhasznált adatok a tranzakció értéke, résztvevői, ideje, és a résztvevők földrajzi helyzete.

2. fejezet

A tranzakciós gráf

A tranzakciós gráf megismerésének a motivációja a Bitcoin hálózat gyakorlati működésének a megértése. Ebben a fejezetben a vizsgált gráfelméleti tulajdonságokat és a valós eseményeket állítom párhuzamba egymással, és keresem az összefüggéseket. A cél, hogy a Bitcoin hálózatát egy már ismert hálózathoz tudjam hasonlítani, feltárni a hasonlóságokat és a különbségeket.

A tranzakciós adatok gyűjtését Juhász Péter, Stéger József, Kondor Dániel és Vattay Gábor végezték a 2018-ban megjelent cikkükhöz [7], ezt bocsájtották a rendelkezésemre. A tranzakciós gráf felépítéséhez és vizsgálatához a Python Pandas és Networkx csomagjait használtam, a 101342 darab tranzakciót egy adattáblába rendeztem.

	0	1	2	3	4	5	6	7	8	9	...
src_user	69545	69545	69545	69545	69545	69545	69545	69545	69545	69545	...
src_latitude	45.7788	45.7788	45.7788	45.7788	45.7788	45.7788	45.7788	45.7788	45.7788	45.7788	...
src_longitude	-119.529	-119.529	-119.529	-119.529	-119.529	-119.529	-119.529	-119.529	-119.529	-119.529	...
src_country	US	US	US	US	US	US	US	US	US	US	...
src_continent	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	...
src_localization	M	M	M	M	M	M	M	M	M	M	...
value	4060861	1343910	2998926	2601546	3389435	607583	5350045	4453744	2515767	4677374	...
timestamp	1385609114	1393207482	1392605247	1388458615	1388288630	1401935884	1380162932	1381890927	1391825956	1382756664	...
txID	28030890	33517326	33062521	30174362	30068344	40128005	24445830	25462184	32482829	26026506	...
dst_user	9070728	9070728	9070728	9070728	9070728	9070728	9070728	9070728	9070728	9070728	...
dst_latitude	42.9864	42.9864	42.9864	42.9864	42.9864	42.9864	42.9864	42.9864	42.9864	42.9864	...
dst_longitude	-78.7279	-78.7279	-78.7279	-78.7279	-78.7279	-78.7279	-78.7279	-78.7279	-78.7279	-78.7279	...
dst_country	US	US	US	US	US	US	US	US	US	US	...
dst_continent	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	...
dst_localization	M	M	M	M	M	M	M	M	M	M	...

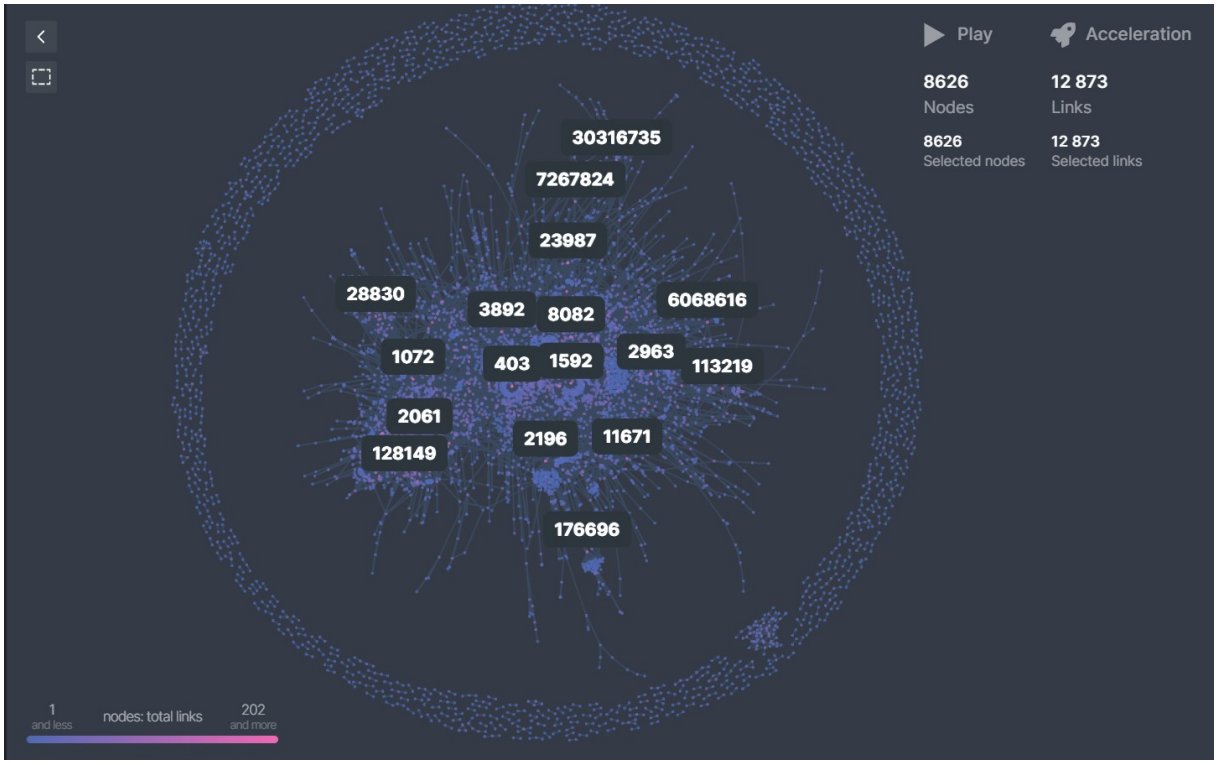
2.1. ábra. A felhasznált adattábla

2.1. A gráf elkészítése

2.1. Definíció. A tranzakciós gráf egy olyan egyszerű gráf, melynek csúcsai a Bitcoin tranzakcióban részt vevő felhasználók, élei pedig a köztük létrejött tranzakciókat reprezentálják. Az irányított és az irányítatlan verziót is használjuk.

A felépített gráfnak összesen 8627 csúcsa és 12874 éle lett a többszörös– és a hurokélek elvétele után. Természetesen a gráf nem összefüggő, 728 összefüggő komponensből áll. A legnagyobb komponens 6882 csúcsot és 10599 élet tartalmaz, a második legnagyobb mindösszesen 61 csúcsból áll. A többi 726 esetben pedig 10-nél kevesebb a csúcsszám.

A gráf struktúrájából jól látszik, hogy már 2013 körül is egy komplex gazdasági ökoszisztémává vált a Bitcoin hálózat. Nem csak egyedi, két fél között menő tranzakciók sokaságáról van szó, hanem megjelennek nagy fokszámú csúcsok is, amelyek a tranzakciók jelentős részének résztvevői. A kapott eredmény összhangban van a tranzakciós gráfról szóló cikkben [9] látottakkal, ahol szintén egy nagy összefüggő rendszer jelenik meg.



2.2. ábra. A tranzakciós gráf

2.2. A gráf tulajdonságai

2.2. Definíció. A $G = (V, E)$ gráf átmérőjének nevezzük és d_G-vel jelöljük a csúcspárok között menő legrövidebb utak közül a leghosszabbnak a hosszát, vagyis

$$d_G = \max_{u,v \in V} \min_{p \in P(u,v)} l(p),$$

ahol $P(u, v)$ az u és v csúcsokat összekötő utak halmaza, $l(p)$ pedig az út hosszát jelöli. Ha a gráf nem összefüggő, akkor az átmérő megegyezés szerint végtelen.

A tranzakciós gráf esetében a legnagyobb összefüggő komponensre az átmérő értéke 18, ami elsőre nagynak tekinthető, viszont elképzelhető, hogy egy láncnak egy tulajdonosa van, aki a címei között mozgatja a bitcoinjait, így a tulajdonosok közötti átmérő érték a valóságban kisebb lehet. A világháló hálózatának átmérője 19-21 közötti érték Barabási Albert-László cikke alapján [1]. A Bitcoin hálózat is nagyban támaszkodik az internetre, ugyanis a kereskedés és az információszerzés is ezen keresztül történik, valamint a szolgáltatások révén a Bitcoin hálózat nagy fokszámú csúcsai az interneten is sok kapcsolódással rendelkezhetnek. Így a kapott érték reálisnak tekinthető.

2.3. Definíció. A $G = (V, E)$ gráf v csúcsára legyen

$$e(v) = \max_{w \in V} \min_{p \in P(v,w)} l(p).$$

Azt mondjuk, hogy egy v csúcs a gráf centruma, ha $\forall w \in V$ csúcsra $e(v) \leq e(w)$ teljesül.

Hasonlóan, mint az előbb, itt is a legnagyobb összefüggő komponenst érdemes vizsgálni. Azt kapjuk, hogy a 403, 1592, 159370, 132580 azonosítóhoz tartozó címek a centrumok, és a hozzájuk tartozó $e(v)$ érték 9. Ezek a címek valószínűleg a rendszer fontos szereplőihez tartoznak, ennek vizsgálata a 3.4 bekezdésben történik.

2.4. Definíció. Egy irányítatlan, egyszerű $G = (V, E)$ gráf sűrűségét D-vel jelöljük, és a

$$D = \frac{2|E|}{|V|(|V| - 1)}$$

képlettel számoljuk ki.

A tranzakciós gráfunk sűrűsége 0,000173, a legnagyobb összefüggő komponensé pedig 0,000447. Ezek az értékek nem meglepő módon kicsik, ugyanis egy felhasználó a legtöbb esetben kevés másikkal hajt végre tranzakciót, leginkább csak néhány szolgáltatóval. A szolgáltatók ugyan sok másik címmel vannak összeköttetésben, de számuk a rendszer résztvevőihez képest alacsony.

2.5. Definíció. Legyen $\lambda_G(v)$ a v csúcsot tartalmazó háromszögek száma. Legyen $\tau_G(v)$ az olyan 2 hosszú utak száma, aminek v a középső csúcsa. Ekkor a v csúcs klaszterezettsége:

$$C_v = \frac{\lambda_G(v)}{\tau_G(v)}$$

Ha a v csúcs foka kevesebb, mint 2, akkor klaszterezettsége 0. A gráf klaszterezettségi együtthatója a csúcsok klaszterezettségének számtani közepe:

$$\frac{1}{|V|} \sum_{v \in V} C_v.$$

A tranzakciós gráf klaszterezettségi együtthatója 0,063, míg a legnagyobb összefüggő komponensének 0,111. Ez azért van így, mert a kis csúcsszámú komponensek jellemzően fák, tehát körmentesek, így háromszög sincs bennük.

A világháló klaszterezettségi együtthatója egy 2017-es cikk [8] szerint 0,1078. Ez megint csak az bizonyítja, hogy a Bitcoin hálózata nagyon hasonló az internetéhez.

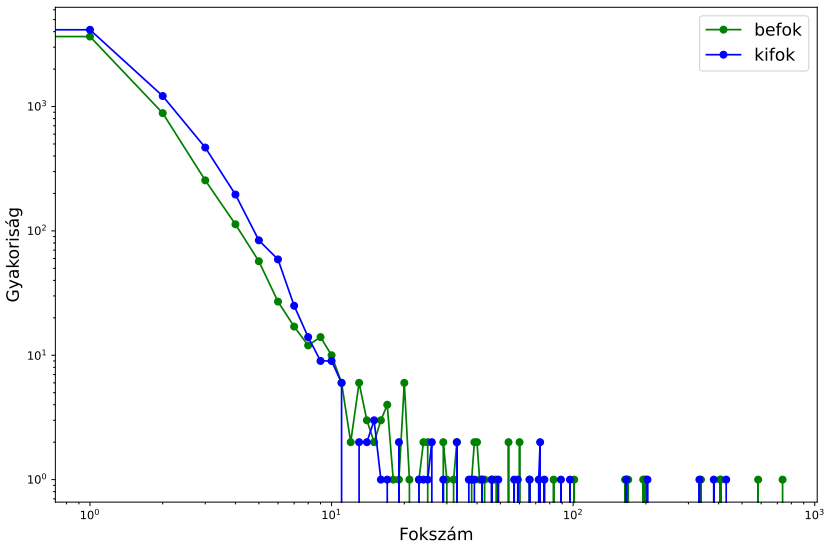
2.6. Definíció. A gráf fokszámeloszlása megadja, hogy az egyes fokszámú csúcsok milyen gyakorisággal fordulnak elő a gráfban, tehát $P(k) = \frac{n_k}{n}$, ahol n_k azon csúcsok száma, melyek fokszáma k, míg n a csúcsok száma. Általában log-log skálán ábrázoljuk (például 2.3 és 2.4 ábra).

2.7. Definíció. Hatványtörvény-eloszlásnak nevezzük az $f : \mathbb{R} \to \mathbb{R}$, $f(x) = cx^{-\alpha}$ alakú függvényeket, ahol $\alpha > 1$, és $c \in \mathbb{R}$. Legfontosabb tulajdonsága a skálainvariancia, aminél ha a függvény argumentumát egy konstanssal (k) megszorozzuk, akkor az eredeti függvényt kapjuk vissza egy másik konstanssal ($k^{-\alpha}$) átskálázva:

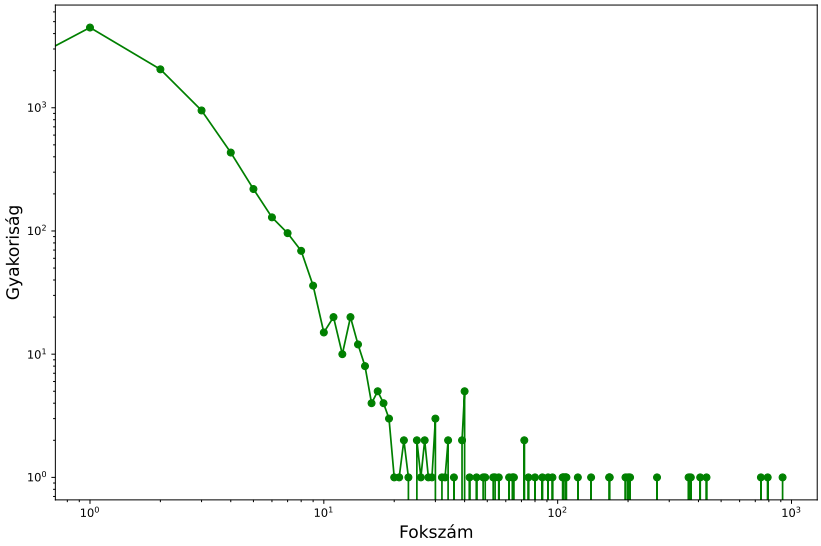
$$f(kx) = c(kx)^{-\alpha} = k^{-\alpha} f(x)$$

A hatványtörvény-eloszlás az egyetlen normalizálható eloszlásfüggvény, amelyre teljesül a skálainvariancia.

Ha a fokszámeloszlás pontjaira egy hatványtörvény-eloszlás illeszkedik, akkor a gráfot skálafüggetlennek nevezzük. Skálafüggetlen hálózat például az internet, a különböző szociális hálózatok, vagy a repülőjáratok hálózata is.



2.3. ábra. Az irányított tranzakciós gráf fokszámeloszlása



2.4. ábra. Az irányítatlan tranzakciós gráf fokszámeloszlása

2.8. Állítás. *A tranzakciós gráf skálafüggetlen.*

Bizonyítás. A bizonyításhoz a *Scale-free networks are rare* című cikkben [4] leírtakat alkalmazzuk. A cikkben szereplő *strongest*, azaz legerősebb feltételek teljesülését vizsgáljuk meg. A kielégíteni kívánt feltételek az alábbiak:

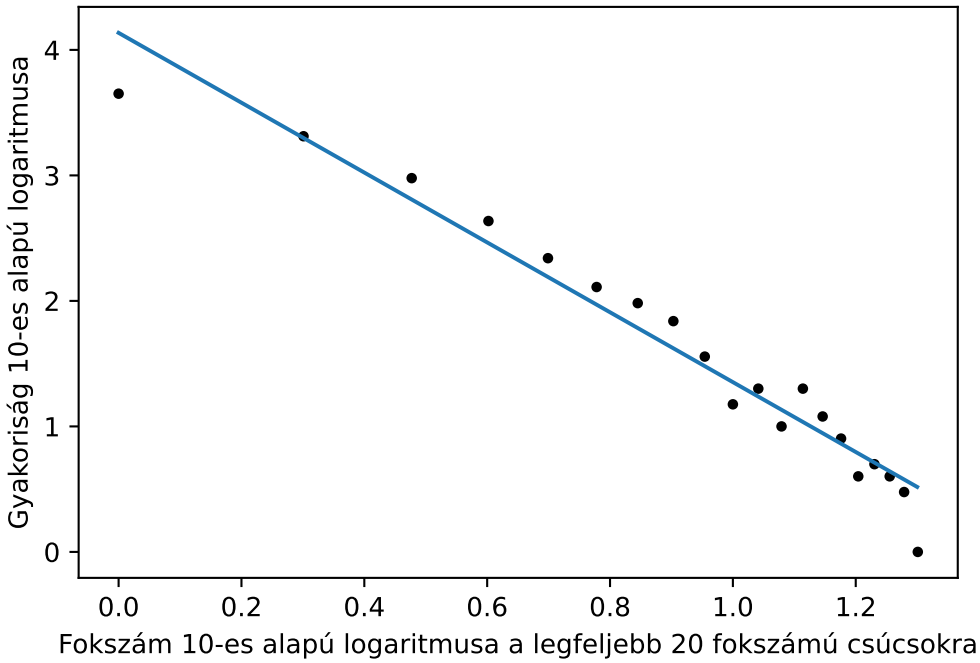
- A fokszámeloszlás egy olyan tartományára, amely a csúcsok legalább 90%-át tartalmazza legyen igaz, hogy hatványtörvény-eloszlást követ
- A hatánytörvény-eloszlás paraméterére $2 < \alpha < 3$ teljesüljön
- A hatványtörvény-eloszlásra illeszkedő tartomány legalább 50 csúcsot tartalmazzon
- A fokszámeloszlás egy olyan tartományára, amely legalább a csúcsok 95%-át tartalmazza legyen igaz, hogy a hatványtörvény-eloszlás illeszkedik a legjobban rá a többi eloszláshoz képest

Ha a fokszámeloszlás valóban hatványtörvény-eloszlást követ, akkor a képének a log-log ábrán egy egyenesnek kell lennie, ugyanis:

$$\log f(x) = \log cx^{-\alpha} = \log c - \alpha \log x,$$

ami a $\log x$-nek lineáris függvénye.

Lineáris regresszióval egyenest illesztünk a fokszámok és a gyakoriság logaritmusára a Python Scypy modul segítségével. A legfeljebb 20 fokszámú csúcsokra illesztünk, melyek száma 8565, tehát a pontok több, mint 99%-a.



2.5. ábra. Lineáris regresszió az adatokra

A lineáris regresszió eredményeként az $\alpha = 2{,}78$ értéket kapjuk, ami megfelel a feltételnek. A standard hibára 0,14-et, míg az R^2-re, ami azt mutatja meg, hogy milyen jól illeszkedik a modell, 0,954-et kapunk. Ezek alapján tehát valóban egy egyenes illeszkedik a pontokra, így a gráf fokszámeloszlása hatványtörvény-eloszlást követ, és semelyik alternatív eloszlás nem illeszkedhet rá jobban. Az összes kritériumot kielégíti a modell, ezzel bebizonyítottuk, hogy a hálózat skálafüggetlen. □

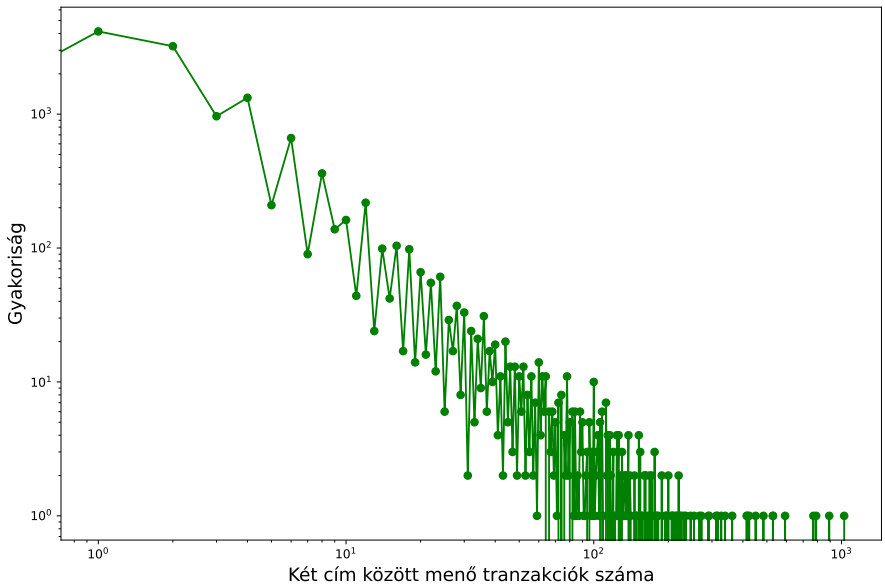
2.9. Definíció. Egy gráf kisvilág-tulajdonságú, ha a gráf méretéhez képest a csúcsok közötti átlagos távolság a csúcsszám logaritmusa, vagy annál kisebb nagyságrendű.

Az internet hálózatára is igaz a kisvilág-tulajdonság. Mivel a tranzakciós gráfunk skálafüggetlen, valamint az átmérője és a klaszterezettségi együtthatója is az internetével azonos nagyságrendű, ezért teljesül rá a kisvilág-tulajdonság.

A következőkben az élek súlyeloszlását fogjuk vizsgálni. Kétféle élsúlyt tudunk alkalmazni az adatok alapján. Az első esetben a két csúcs között menő tranzakciók számát írjuk rá az élekre. Ezt szintén log-log skálán ábrázoljuk (2.6 ábra).

A fokszámeloszlással ellentétben ez az eloszlás jóval zajosabb. Az eloszlás farok tartományába az élek jelentős része tartozik. A legfeljebb 10 súlyú élek száma 11275, a két csúcs között menő tranzakciók legnagyobb száma pedig 1025. Több, mint 100 esetben legalább 100 tranzakció jött létre két cím között.

Mindezekből arra lehet következtetni, hogy a tranzakciók jelentős részében a különböző szolgáltatók vesznek részt, ugyanis ők képesek ennyi tranzakciót lebonyolítani, ilyen lehet például a Satoshi Dice is. Ezek a nagy súlyú élek akár két szolgáltató között is futhatnak, például egy Bitcoin tőzsde és egy Wallet szolgáltató között.

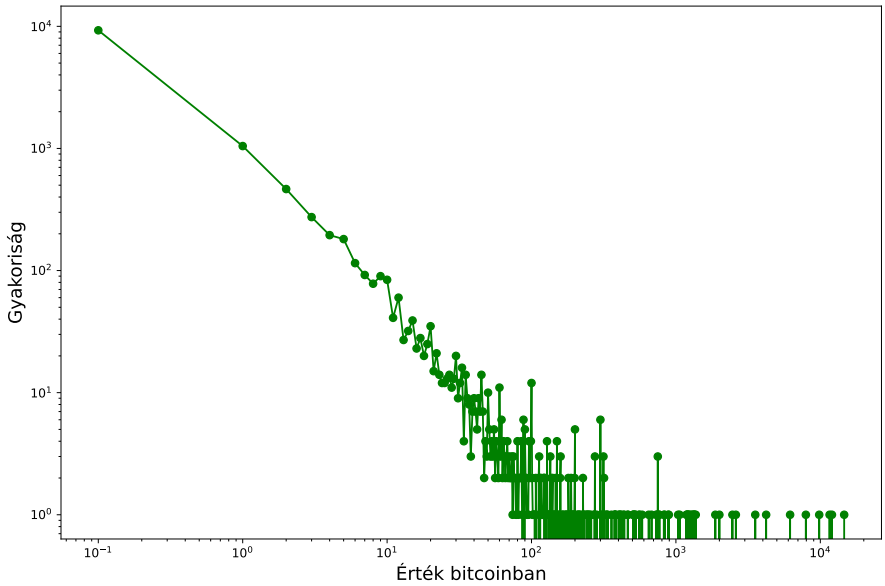


2.6. ábra. Tranzakciók számának eloszlása

A második esetben a két cím között menő tranzakciók bitcoin mennyiségének az összegét írjuk az élekre, és szintén log-log skálán ábrázoljuk (2.7 ábra).

Az ábra elkészítésének érdekében a tranzakcióban szereplő bitcoin mennyiségnek az alsó egészrészét vesszük, az 1-nél kisebb esetek pedig a 10^{-1}-nél jelennek meg.

Megfigyelhetjük, hogy közel 7000 esetben a címek között 1-nél kevesebb bitcoin cserél gazdát. Itt is nagyon jelentős a farok tartományba tartozók száma, ami az előzőekben látottak szerint a rendszeres bitcoin küldésnek köszönhetően is lehetséges. A legnagyobb összértékek 10000 bitcoin felett vannak, amik valószínűleg szintén a szolgáltatókhoz köthetőek.



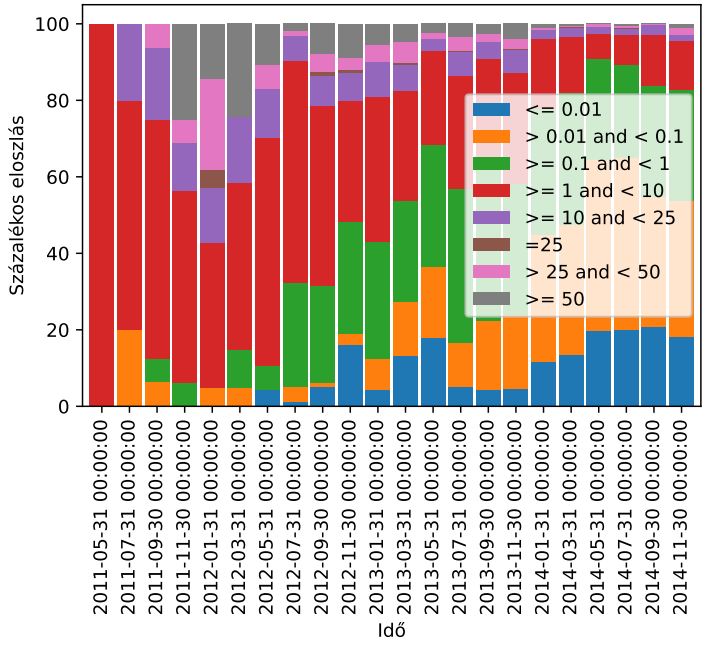
2.7. ábra. Értékek eloszlása

2.3. További tranzakciókhoz köthető vizsgálatok

Érdemes megvizsgálni a tranzakcióban szereplő bitcoinok nagyságának eloszlását az idő függvényében is (2.8 ábra).

Azt vehetjük észre, hogy az idő teltével egyre dominánsabbá válnak az 1 bitcoin alatti tranzakciók. Ez a tendecia két tényező miatt alakult így. Az egyik az árfolyam növekedéséhez kötődik. Ebben az időszakban volt az első jelentős kiugrása az árfolyamnak (1.4 ábra), ami hozzájárult a kisebb tranzakciók megjelenéséhez, ugyanis azonos értékű pénz küldéséhez már kevesebb bitcoinra volt szükség.

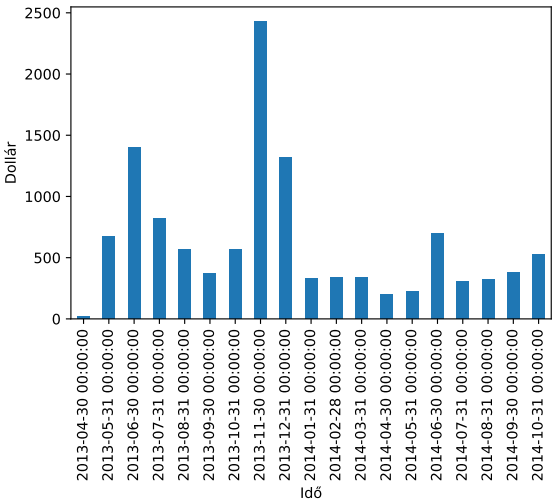
A másik tényező pedig a Satoshi Dice megjelenése volt a [9] cikkben leírtak alapján. Ez a szerencsejáték egy digitális sorsjegyhez hasonlítható. Különböző Bitcoin címek voltak megadva a játék megnyerésének valószínűségével (0,0015%-tól 97%-ig), valamint egy oddsszal, ami megadta, hogy a tétünk hányszorosát nyerhetjük meg. A játékhoz szükséges bitcoin elküldése után egy automatizált rendszer visszaküldte a megadott oddsszal megszorzott bitcoint győzelem esetén, ellenkező esetben pedig 1 satoshit. A kimenetel eldöntését egy algortimus számolta ki a tranzakció hashének és egy random számnak a segítségével. Jellemzően alacsony téttel játszottak, a fogadások több, mint ötöde a minimum téttel (0,01 BTC) történt. A tranzakciók meghatározó részét tették ki, ugyanis egy nap akár 30000 játékot is játszottak ezen keresztül.



2.8. ábra. Tranzakció értékeinek eloszlása az időben

Egy másik vizsgálat a tranzakciók dollárban mért nagyságának időbeli eloszlása. Ennek meghatározására a Python Cryptocmd modulját használtam, amelyben 2013. április 28-ai volt az első árfolyamadat. A tranzakciók időbélyegének segítségével meghatároztam a kívánt értékeket. A 2.9 ábrán az értékek átlagának havi eloszlása látható.

A kiugró érték egyértelműen a 2013-as év végi árfolyamnövekedésnek köszönhető. A magas értékből az látszik, hogy sokan már ekkor befektetési célból használták a bitcoint, és feltehetőleg ebben az időben realizálták a profitjukat.



2.9. ábra. Tranzakció értékeinek eloszlása az időben

2.4. Összegzés

A fejezetben megvizsgált gráfelméleti tulajdonságok alapján azt mondhatjuk, hogy a Bitcoin hálózata sok szempontból hasonlít az internet hálózatára. Kiemelendő a skálafüggetlenség és a kisvilág-tulajdonság, amely a szociális hálózatokra is jellemző. Az eredmények alapján nagy jelentőséggel bír a tranzakciós gráfra az árfolyam mozgása és a gazdasági szereplők. Akár egyetlen új szolgáltató (Satoshi Dice) is képes volt szignifikáns változásokat előidézni a tranzakciók számát és a küldött bitcoinok mennyiségét illetően is. A folyamatokat tehát sok tényező alakítja, de az elmondható, hogy a rendszer rövid időn belül egy meghatározó, komplex gazdasági ökoszisztémává vált.

3. fejezet

A tranzakciók távolságának skálázódása

A vizsgálat motivációját a 2006-ban írt *The scaling laws of human travel* című cikk képzi [3]. Az ebben leírtakat alapul véve elemzem a készpénz és a digitális pénz, azaz a bitcoin, térbeli és időbeli mozgását. A célom, hogy feltárjam a hasonlóságokat és különbségeket a különböző pénzáramlások között. A fejezetben továbbá statisztikai szempontok szerint is vizsgálom a bitcoin tranzakciók földrajzi eloszlását, valamint kitérek bizonyos Bitcoin címek deanonimizálására is.

3.1. A motivációt adó cikk leírása

A cikk célja az emberek utazásainak, mozgásának eloszlását leíró modell megalkotása, mert ezzel lehet modellelzni például egy vírus terjedését is. A kutatás alapját egy online adatbázis adta, a *wheresgeorge.com*, ahol amerikai dollár bankjegyek mozgását lehet nyomon követni azáltal, hogy az emberek a megkapott bankjegy helyzetét frissítik az oldalon az egyedi kódjuk alapján. A feltételezés, hogy a készpénz mozgása jól leírja az ember mozgását a készpénz fizikai mivoltából fakad, ugyanis ahhoz, hogy egyik helyről a másikra eljusson emberi tényező közbeiktatása szükséges. Összesen 464670 bankjegy 1033195 földrajzi helyzetének és két mozgás között eltelt időnek felhasználásával történt a vizsgálat.

3.1. Definíció. A Lévy-eloszlás egy abszolút folytonos eloszlás, sűrűségfüggvénye:

$$f(x) = \sqrt{\frac{c}{2\pi}} \frac{e^{-\frac{c}{2(x-\mu)}}}{(x-\mu)^{\frac{3}{2}}},$$

ahol $x > \mu$, és μ a helyparaméter, $c > 0$ a skálaparaméter. Ha $x \leq \mu$, akkor $f(x) = 0$. Ez egy hosszú farkú eloszlás, tehát a kiugróan nagy értékek is viszonylag nagy valószínűséggel fordulnak elő. A várható értéke és a szórása egyaránt végtelen.

3.2. Definíció. A Lévy-repülés egy olyan véletlen bolyongás, melynél a lépések függetlenek, és a lépések hossza nem állandó, hanem Lévy-eloszlást követ. Ez azt jelenti tehát,

hogy gyakoriak a kis lépések, de időnként egy-egy nagyobb ugrás is bekövetkezik. Rengeteg különböző területen fordul elő, pédául a pénzügyi matematikában és a fizikában is.

A Lévy-repülés jó leírása lehet egy ember mozgásának, ugyanis sokszor kicsit lépünk, például ha boltba vagy munkahelyre megyünk, aztán ritkán előfordul egy nagy ugrás, ilyen lehet egy nyaralás.

Ha a bankjegyek mozgására is Lévy-repülést feltételezünk, mert ezzel akarjuk leírni az emberek mozgását, akkor körülbelül 68 nap alatt a bankjegyek egy egyenletes eloszlást érnének el, tehát bárhol lehetnek. A mérések viszont nem ezt mutatták, ugyanis közel egy év elteltével is a bankjegyek mindössze 23,6%-a jutott 800 kilométernél messzebbre, míg 19,1%-uk egy 50 kilométer sugarú körben maradt. Tehát az egyszerű Lévy-repülés nem teljesen jól írja le a mozgást, kiegészítésre szorul.

A cikk két alternatív magyarázattal szolgál arra, hogy a bankjegyek miért mozognak sokáig csak egy kis sugarú körben. Az egyik a rendszer térbeli inhomogenitása, jellemzően a nagyvárosokat ritkábban hagyják el az ott élők, mint a többi kisebb települést a lakói. A másik indok a két mozgás között eltelt várakozási időhöz kapcsolódik, ami sok esetben elég nagy volt.

A probléma megoldásának érdekében először azt vizsgálták meg, hogy egy bankjegy az idő függvényében milyen valószínűséggel marad a kezdeti pozíciójának 20 kilométeres körzetében. A vizsgálatot három csoportra is elvégezték, ezek a nagy–, közép–, és kisvárosok voltak. Mindegyik csoport esetén aszimptotikus viselkedést figyeltek meg, $P(t) \sim At^{-\eta}$, ugyanazzal az $\eta = 0{,}6 \pm 0{,}03$ kitevővel. Ebből az következik, hogy a várakozási idő és a térbeli szóródás karakterisztikája univerzális. Ugyanerre az egyszerű Lévy-repülést feltételezve $\eta \approx 3{,}33$ jönne ki, ami megint csak azt mutatja, hogy módosítás szükséges.

3.3. Definíció. A folytonos idejű véletlen bolyongás a véletlen bolyongások egy általánosítása, amelynél a két lépés között eltelt várakozási idő is egy valószínűségi változó. Formalizálva, ha $X(t)$ egy véletlen bolyongás [5]:

$$X(t) = X_0 + \sum_{i=1}^{N(t)} \Delta X_i,$$

ahol ΔX_i független, azonos eloszlású valószínűségi változók, $N(t)$ pedig a lépések száma a $(0, t)$ intervallumon, akkor annak a valószínűsége, hogy a bolyongás az k értéket veszi fel a t időben:

$$P(k, t) = \sum_{n=0}^{\infty} P_l(n, t) P_n(k),$$

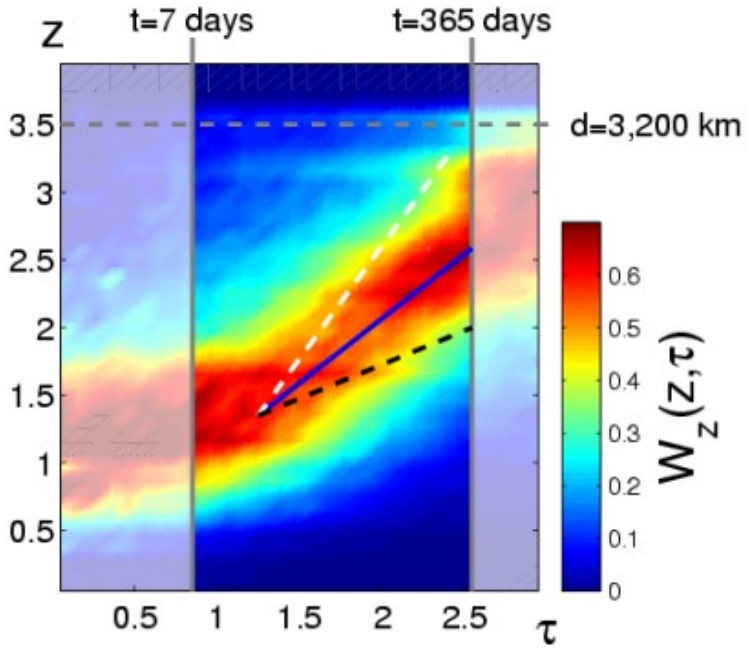
ahol $P_l(n,t)$ annak a valószínűsége, hogy n lépés történt t idő alatt, míg $P_n(k)$ annak a valószínűsége, hogy a bolyongás az k értéket veszi fel n lépés után.

A cikkben a modell hibáinak kiküszöbölésére folytonos idejű véletlen bolyongást alkalmaznak a bankjegyek elmozdulásának nagyságára és a várakozási idő hosszára. Mindkettőnek az eloszlására azt feltételezték, hogy hatványtörvény-eloszlást követnek különböző paraméterekkel. A kapott eredmény szerint a valószínűségi eloszlása annak, hogy t idő alatt legfeljebb r távolságra jutott:

$$W_r(r,t) = t^{-1/\mu} L\left(\frac{r}{t^{1/\mu}}\right),$$

ahol L egy univerzális skálafüggvény, és a modell alapján várhatóan $\mu = 0{,}98 \pm 0{,}08$.

Az adathalmaz egészére alkalmazva a modellt $\mu = 1{,}05 \pm 0{,}02$ eredmény jött ki, ami jól illeszkedik az eddig leírtakhoz. Ezt egy $z = \log_{10} r$, $\tau = \log_{10} t$ átskálázással és a hozzá tartozó $W_z(z,\tau)$ eloszlással ábrázolták (3.1 ábra).



3.1. ábra. A bankjegyek elmozdulása az idő függvényében [3]

A kék vonal jelenti a modell alapján megkapott $\mu = 1{,}05 \pm 0{,}02$-es értéket, míg a fehér szaggatott vonal az egyszerű Lévy-repülést. Az ábra jól mutatja, hogy a bankjegyek sokáig mozognak egy kis sugarú környezetben, majd egy nagyobb ugrással kerülnek messzebb eredeti pozíciójuktól. A modell tehát jól alkalmazható a bankjegyek mozgásának leírására.

3.2. Az adatokat biztosító cikk leírása

A felhasznált adatok az *A Bayesian approach to identify Bitcoin users* című cikk [7] eredményei. Részletesebben megvizsgáljuk hogyan is zajlott az adatgyűjtés és elemzés folyamata.

A Bitcoin blokkláncáról szerzett adatok és egy valószínűségszámítási modell segítségével lehetőség nyílik a tranzakciót kezdeményező felhasználók azonosítására. A vizsgálat két fontos lépést tartalmaz. Először is, minden egyes tranzakció esetében meghatározzák annak a valószínűségét, hogy egy adott felhasználó, az IP-címe alapján, létrehozta-e azt. Feltételezve, hogy a tranzakció létrehozója birtokolja azokat a Bitcoin címeket, amelyekről bitcoint küldenek, ez a lépés lehetséges IP-cím – Bitcoin cím párosításokat eredményez. Ezután a legvalószínűbb IP-cím – Bitcoin cím párosítások azonosítása az előző lépésben összeállított párosítási listában szereplő valószínűségek kombinálásával történik. Az azonosítás módszere továbbfejleszthető a tranzakciós hálózat alapján nagy valószínűséggel ugyanahhoz a felhasználóhoz tartozó Bitcoin címek csoportosításával. Végül pedig a földrajzi helymeghatározás az IP-címek segítségével történik.

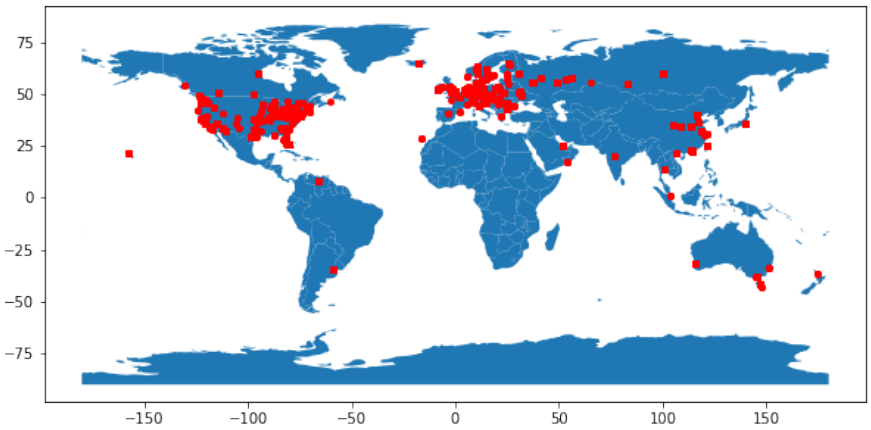
Ennek eredményeként 22363 felhasználót sikerült azonosítani, és összesen 1797 IP-címet rendeltek hozzájuk. A kiegyensúlyozatlanságot három IP-cím okozza, amelyekhez 20680 felhasználó van hozzárendelve. Ezek az IP-címek valószínűleg a Wallet szolgáltatókhoz tartoznak, ugyanis ők képesek ennyi Bitcoin címet egyszerre kezelni.

3.3. A tranzakciók távolságának vizsgálata

Az elemezés a tranzakciók tulajdoságait tartalamzó adattábla (2.1 ábra) segítségével történt. A vizsgálatok jelentős részét a távolság és a tranzakció többi paraméterének függvényében végeztem Python segítségével.

3.3.1. A felhasználók földrajzi eloszlása

A Python Geopandas modulját felhasználva a 3.2 ábrán piros négyzettel jelölve láthatjuk a felhasználókat. Azt vehetjük észre, hogy az Egyesült Államok és Európa területére koncentrálódik a felhasználók meghatározó része. Megemlítendő még Kína szerepe, ahol valószínűleg bányász tevékenységet folytattak. Érdekes módon Közép-Amerikából és Afrikából nincsenek tranzakciós adatok. Az elmúlt 10 évben viszont tudjuk, hogy ez változott, mivel ezen kontinensek egyes országaiban a rossz gazdasági helyzet, magas infláció miatt sokan döntöttek a Bitcoinba való befektetés mellett. Összességében tehát elmondható, hogy 2013 körül a fejlett világ országait sűrűn behálózták, míg a fejlődő országokban kevésbé jelentek meg Bitcoin felhasználók.



3.2. ábra. A felhasználók földrajzi eloszlása

3.3.2. A távolság meghatározása a haversine formulával

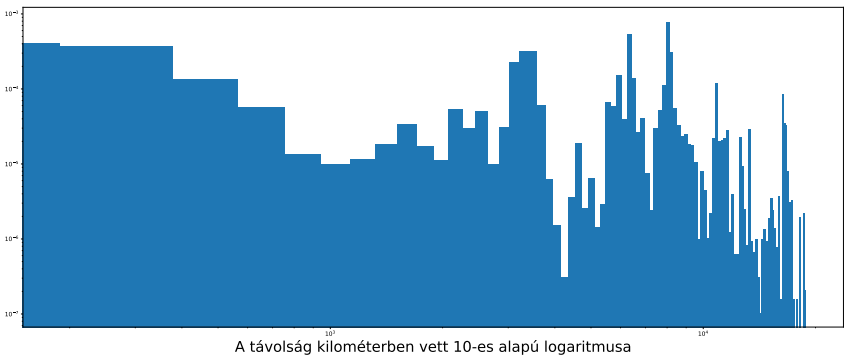
Mivel a rendelkezésemre álló adatban a küldő és a címzett szélességi és hosszúsági koordinátái szerepelnek, így a tranzakció távolságának kiszámításához a *haversine formulát* használtam. A haversine formula kiszámolja a gömbfelület két pontja között menő legrövidebb út hosszát, ha a pontok szélességi és hosszúsági koordinátákkal vannak megadva [13]. A haversine formula implementáláshoz a GeeksForGeeks weboldalon leírtakat vettem alapul.

```python
import numpy as np
def Haversine(lat1,lon1,lat2,lon2, **kwarg):
    """
    This uses the 'haversine' formula to calculate the great-circle distance between two points - that is,
    the shortest distance over the earth's surface - giving an 'as-the-crow-flies' distance between the points
    (ignoring any hills they fly over, of course!).
    Haversine
    formula:    a = sin²(Δφ/2) + cos φ1 · cos φ2 · sin²(Δλ/2)
    c = 2 · atan2( √a, √(1−a) )
    d = R · c
    where   φ is latitude, λ is longitude, R is earth's radius (mean radius = 6,371km);
    note that angles need to be in radians to pass to trig functions!
    """
    R = 6371.0088
    lat1,lon1,lat2,lon2 = map(np.radians, [lat1,lon1,lat2,lon2])

    dlat = lat2 - lat1
    dlon = lon2 - lon1
    a = np.sin(dlat/2)**2 + np.cos(lat1) * np.cos(lat2) * np.sin(dlon/2) **2
    c = 2 * np.arctan2(a**0.5, (1-a)**0.5)
    d = R * c

    return round(d,1)
```

3.3. ábra. A haversine formula implementálása [12]

3.3.3. A távolság különböző eloszlásai

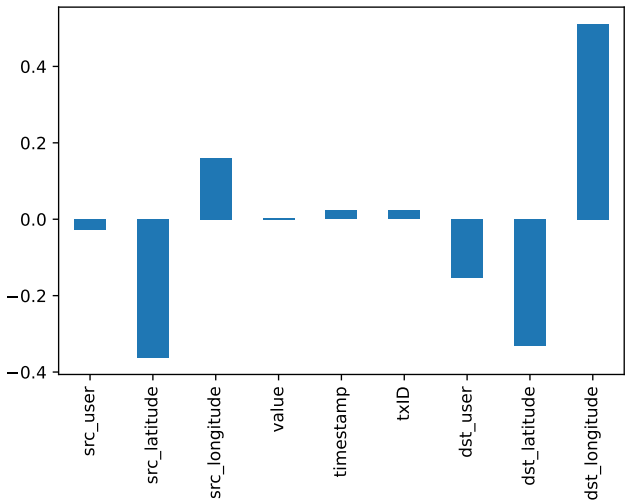
Először vizsgáljuk meg a távolság eloszlásának hisztogramját log-log skálán ábrázolva (3.4 ábra).



3.4. ábra. A távolság eloszlásának hisztogramja

Azt vehetjük észre, hogy nagyon jelentős az 5000 kilométernél is nagyobb távolsággal rendelkező tranzakciók aránya, ami ellentétes viselkedést mutat a készpénznél látottakkal szemben. Tehát a bitcoin digitális létezéséből kifolyólag leküzdi a fizikai akadályokat, és igen nagy távolságokat képes megtenni. Ismert eloszlás nem illeszkedik rá, ezért további vizsgálatok szükségesek az eloszlás jobb megértéséhez.

Fontos megnézni, hogy a tranzakció mely paramétereivel korrelál a távolság (3.5 ábra).

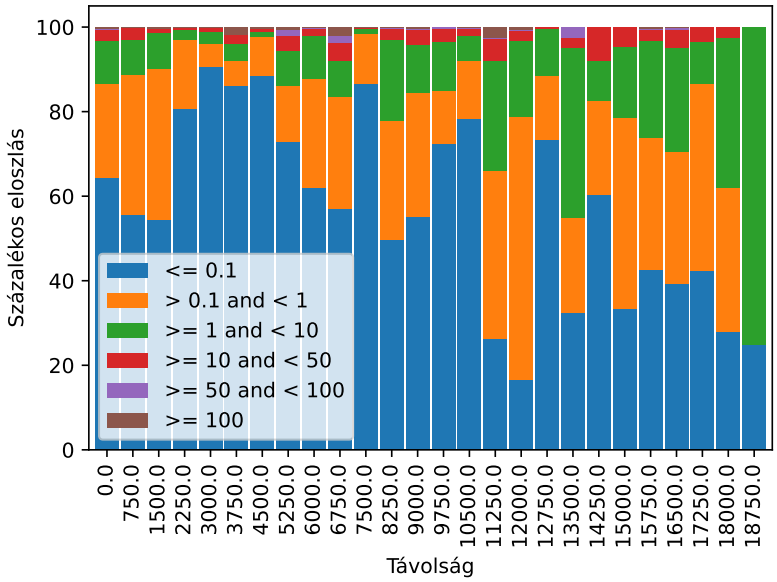


3.5. ábra. A távolság korrelációja a tranzakció többi paraméterével

A szélességi és hosszúsági fokokból korrelációjából az látszik, hogy ha a tranzakció valamelyik résztvevője minél keletebbre és délebbre helyezkedik el, annál nagyobb lesz a távolság. Ez arra utal, hogy az egyik résztvevőnek európainak vagy ázsiainak kell lennie ahhoz, hogy igazán nagy távolság alakuljon ki, mert a tranzakciók jelentős részének egyik résztvevője amerikai.

Meglepő módon a távolság és az érték között csupán 0,0026 a korreláció, tehát korrelálatlannak tekinthetjük őket. Mivel a készpénz esetén csak az 1 dolláros bankjegyek mozgását figyelték meg, ezért pontos összehasonlítást nem tudunk végezni. Az intuició alapján viszont azt gondolhatjuk, hogy sok készpénzt veszélyesebb lehet nagy távolságra eljuttatni, mert könnyebb elveszíteni, míg a Bitcoin digitális hálózatán ez a veszély nem áll fenn.

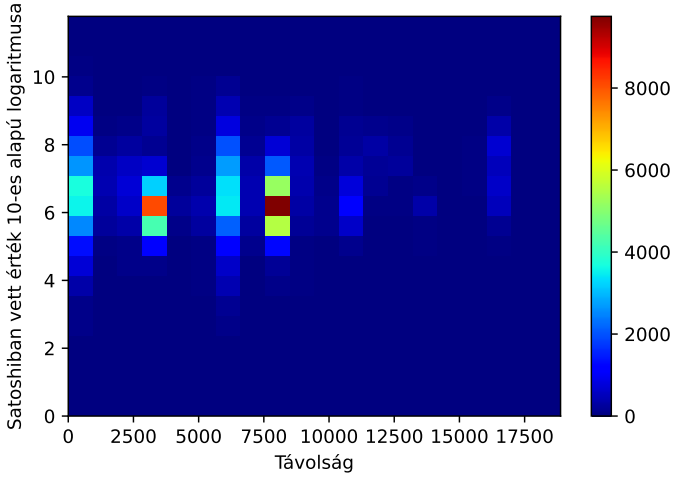
Az előzőleg leírtak ellenőrzésére vizsgáljuk meg a tranzakciók értékének százalékos eloszlását a távolság függvényében (3.6 ábra).



3.6. ábra. A tranzakciók értékének eloszlása a távolság függvényében

Valóban azt láthatjuk, hogy gyakorlatilag minden távolság esetén az 1 bitcoinnál kisebb nagyságú tranzakciók dominálnak. Ez egybecseng a korrelációnál látottakkal. A legnagyobb távolságoknál figyelhető meg egy kicsi értéknövekedés, de mivel az ilyen tranzakciók száma nem összemérhető a többivel, ezért van csak minimális pozitív korreláció. Érdekesség, hogy a 2.8 ábrával ellentétben, ami a tranzakciók értékeinek százalékos eloszlását az idő függvényében ábrázolja, sokkal kevésbé dominálnak az olyan tranzakciók, melyek értéke 1 és 10 bitcoin közé esik. Ez azt jelenti, hogy ha egy érték kategóriába sok tranzakció esik, akkor azok is egyenletesen oszlanak el a távolság függvényében.

Az előző 3.6 ábrán csak a távolság tartományokon belüli relatív eloszlását láttuk az értékeknek, de azt nem, hogy ez valójában hány tranzakciót jelent. Ezért vizsgáljuk meg a 3.7 ábrát is. Ezen az értékek nagyságának mennyiségi eloszlását látjuk a távolság függvényében. Mivel 10^8 satoshi egyenlő 1 bitcoinnal, ezért az y tengelyen a 6-os a 0,01 míg a 8-as az 1 bitcoint jelenti.



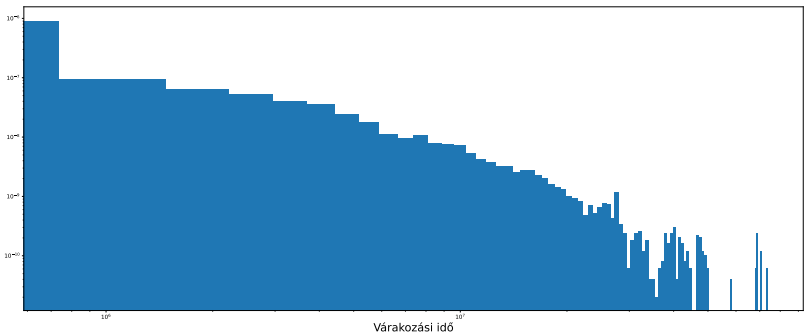
3.7. ábra. A tranzakciók értékének eloszlása a távolság függvényében

Szembetűnően látszik, hogy 3000 és 8000 kilométer körüli távolság tartományban jelentős a tranzakciók száma, ezeken belül is a 0,01 és 1 bitcoin közötti értékűeké. Ez meghatározó gazdasági kapcsolatra utal, ami valószínűleg szolgáltatók között jött létre. A 69545 és a 403 azonosítójú felhasználóknak is több ezer tranzakciója ezekbe a távolság tartományokba esik. Ezen címek deanonimizálása a 3.4 bekezdésben történik.

A következőkben egy kettő hosszú tranzakciós lánc vizsgálatával foglalkozunk. Azzal a heurisztikával élünk, hogy egy bitcoin következő tranzakciója akkor következik be, amikor a címzett küldővé válik. Ez természetesen nem vezet tűpontos eredményre, mivel nem biztos, hogy valóban ugyanazokat a bitcoinokat költik el megint a felhasználók, de ugyanakkor információnk lesz a várakozási időről, és a kettő hosszú láncok távolságát is meg tudjuk határozni.

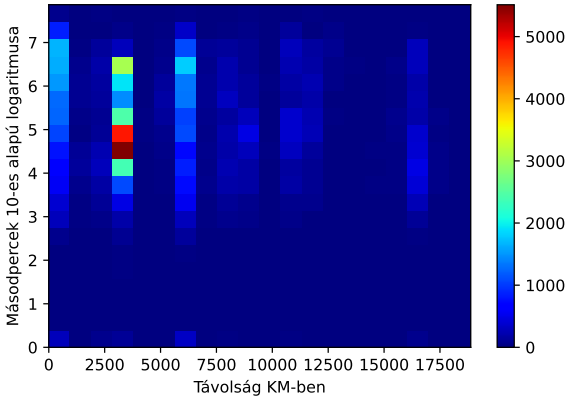
Az adattáblában a tranzakciókat az időbélyegük szerint növekvő sorrendbe rendezzük, majd minden címzett esetén megnézzük, hogy mikor szerepel először küldőként. Ekkor vesszük az időbélyegek különbségét és hozzárendeljük a tranzakció távolságát is. A láncok távolságát pedig az első tranzakció küldőjének és a második tranzakció címzettjének koordinátáival számoljuk ki a haversine formula segítségével.

Vizsgáljuk meg először a várakozási idő eloszlását log-log skálán (3.8 ábra). A 10^6 másodperc felel meg körülbelül egy napnak, míg a 10^7 másodperc nagyjából két hét. A bitcoinok jelentős része rövid időn belűl gazdát cserél, ami egyrészt a Satoshi Dice következménye is lehet, mivel ugyanazzal a bitcoinnal sok játékot játszhattak. Másrészről valószínűleg a bitcoinok egy jelentős része egyáltalán nem is mozog, mivel nagyrészük egyes statisztikák szerint már elveszett elfelejtett privát kulcsok miatt. Így az mondható, hogy amelyik bitcoinok mozognak, azok gyakran kerülnek tranzakcióba.



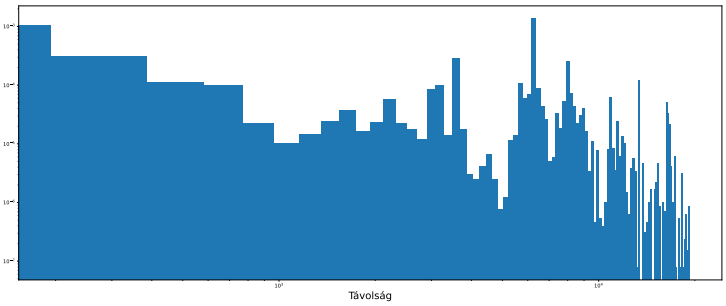
3.8. ábra. A várakozási idő eloszlása

Nézzük meg, hogy a távolság függvényében hogyan oszlanak el a várakozási idők (3.9 ábra). Itt is azt láthatjuk, hogy a bitcoinok kardinális részénél két hétnél kevesebb idő kell, hogy újra tranzakcióban szerepeljen. A 3000 kilométeres tartományban rajzolódik ki meghatározó számú tranzakció, csak úgy, mint 3.7 ábra esetén. Ez valószínűleg ugyanahhoz a gazdasági kapcsolathoz fűződhet. A korrelációt megvizsgálva $-0{,}029$ jön ki, ami korrelátlannak tekinthető, tehát a várakozási idő nem befolyásolja a tranzakció távolságát.



3.9. ábra. A várakozási idő eloszlása a távolság függvényében

Utoljára pedig nézzük meg a kettő hosszú lánc első és harmadik résztvevőjének távolságát (3.10 ábra). Itt egy nagyon hasonló eloszlást láthatunk, mint a 3.4 ábrán. Tehát azt mondhatjuk, hogy a bitcoinok két lépés után is nagyjából ugyanolyan messze vannak a kezdeti pozíciójuktól, mint egy lépés után.

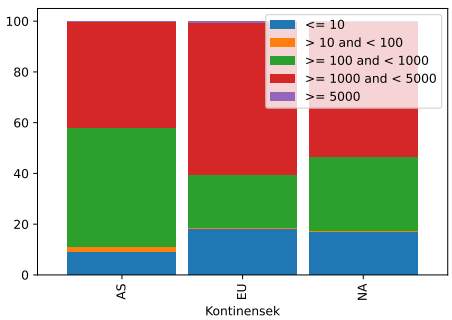


3.10. ábra. A várakozási idő eloszlása a távolság függvényében

3.3.4. Kontinenseken belüli eloszlás vizsgálata szórásanalízissel

A szórásanalízis egy olyan hipotézisvizsgálati eljárás, amely során ugyanazt a mennyiséget vizsgáljuk bizonyos tulajdonságok szerint különböző csoportokba sorolva. A cél, hogy eldöntsük, hogy ennek a mennyiségnek az egyes csoportora jellemző eloszlásának ugyanaz-e a várható értéke. Másképpen megfogalmazva, igaz-e, hogy a vizsgált mennyiség várható értékére nincs hatása annak, hogy a megfigyelés melyik csoportból származik. Ez a kétmintás Student-féle t-próba egy többmintás általánosítása [14].

A vizsgálat a Python Scipy.stats moduljának segítségével zajlott. Először is a vizsgált mennyiség a tranzakciók távolsága lesz, a csoportok pedig az egyes kontinenseken belüli tranzakciók. Nézzük meg ezek eloszlását (3.11 ábra). Az *AS, EU* és *NA* rövidítések rendre Ázsia, Európa és Észak-Amerika kontinenseket jelentik.

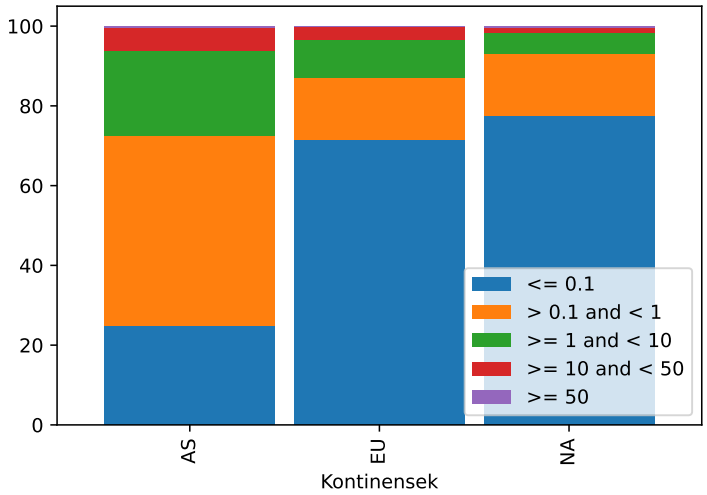


3.11. ábra. A tranzakciók távolságának eloszlása kontinensek szerint

Elsőre meglepő lehet, de mindhárom esetben hasonló eloszlást tapasztalunk. Mivel az észak-amerikai tranzakciók nagyrésze az Egyesült Államokon belül zajlott, melynek mérete nagyjából Európáéval azonos, így valóban hasonló eloszlásra számíthatunk. Ázsia esetében pedig a jelentős rész Kínán belül bonyolódott le, melynek területe szintén hasonló nagyságú Európáéhoz. Európa esetén megjegyzendő, hogy sok tranzakcióban oroszországi címek szerepelnek, melyek a nagy távolságokat okozzák.

Ennek ellenére szórásanalízist alkalmazva a p-értékre $2 \cdot 10^{-24}$ jön ki, ami azt jelenti, hogy legalább egy csoportban szignifikáns eltérés van a várható értéket illetően. Ez azért is lehetséges, mert a 3.11 ábrán az 1000-5000 kilométeres tartományba eső tranzakcióknál nagy szóródás figyelhető meg a különböző kontinenseken belül, ami kihatással van a várható értékre. Végső soron egy ellentétes eredményt kapunk, tehát a kontinensen belüli tranzakciók esetén a távolság várható értéke eltérő.

A második vizsgálatunk a kontinensen belüli tranzakciók értékére irányul. Nézzük meg itt is először az eloszlást (3.12 ábra).



3.12. ábra. A tranzakciók értékének eloszlása kontinensek szerint

Európa és Észak-Amerika esetében hasonló eloszlást tapasztalunk, míg Ázsiában a 0,1 bitcoinnál kisebb értékű tranzakciók aránya kisebb a többi kontinenshez képest. Alkalmazva itt is a szórásanalízist a p-értékre 0,76 jön ki. Eszerint nincs szignifikáns eltérés a várható értékben a különböző kontinenseknél.

3.4. Bitcoin felhasználók deanonimizálása

A felhasznált adattáblában nem konkrét Bitcoin címek és tranzakciós azonosítók szerepeltek, hanem egy új elkódolása ezeknek, ami megnehezítette a keresést. Ezért a tranzakciókat az időbélyegük segítségével tudjuk a blockchain.com weboldalon visszakeresni, hogy melyik blokkba kerültek be, majd a blokkon belül már a tranzakcióknál meg lehet keresni az érték alapján a vizsgált tranzakciót.

A keresést először a nagy értékű tranzakciókkal kezdjük. A 25 legnagyobb értékű tranzakciót vizsgáljuk, ezek mindegyikében több, mint 1000 bitcoint küldtek. A legtöbb címhez nem volt található semmilyen információ az interneten, pár esetben viszont Mircea Popescuhoz kapcsolódó oldalak jelentek meg. Mircea Popescu a valaha élt legtöbb bitcoinnal rendelkező személy volt, az MPEx kereskedőfelület és a Satoshi Dice létrehozója. A talált cím az *1JPvucRfu3ZzEvfBUQTJwsxMrZjeTqD6zR*, ami tehát Popescuhoz tartozott, és bizonyos befektetéseket bonyolított ezen keresztül. Az adattáblában tehát a 41025-ös azonosító tartozik hozzá, és az Egyesült Államokban, Buffaloban található.

Második lehetőségként a legalább 500 tranzakcióval rendelkező címeket próbáljuk azonosítani. Sikerült a CaVirtEx, McxNOW, FYB-SG, valamint az Okcoin tőzsdékhez tartozó címeket felderíteni a bitinfocharts.com weboldal segítségével. Ezekhez rendre a 3574, 1592, 11856, és a 2285-ös azonosítók tartoznak. Sikerült továbbá az egyik legnagyobb Mining poolt, az F2Pool-t is deanonimizálni az 114426 címke alatt. A szerencsejátékhoz köthető szolgáltatások közül a BetVip a harmadik legtöbb tranzakcióval rendelkező 403-as címhez kötődik. Végül pedig a PandoraOpenMarket illegális termékekkel is kereskedő oldalt sikerült a 3892-es címke alatt azonosítani. Ezek eloszlása nagyon változatos, ugyanis Amerikában, Kínában és Ausztráliában helyezkednek el. Többek között ez is okozhatja a nagy távolságok kialakulását.

Sok esetben a keresés eredményeképp wallet azonosítók jöttek ki, amik nem voltak felderítve. A két legtöbb tranzakcióval rendelkező cím esetében is ez volt a helyzet. A 69545-ös cím a tranzakciók mintázata alapján egyértelműen a Satoshi Dice játékhoz köthető, ugyanis gyakran egy blokkon belül a bitcoinok több tranzakcióban is szerepeltek. A 2061-es cím tartozhatna a Mt. Gox tőzsdéhez, ami nagyon jelentős volt ebben az időben, viszont a tranzakciók eloszlása inkább egy Mining pooléhoz hasonló.

3.5. Összegzés

Ebben a fejezetben összehasonlítottuk a készpénz és a bitcoin földrajzi mozgásának viselkedését. Azt tapasztalhattuk, hogy a Bitcoin digitális rendszerén keresztül sokkal nagyobb távolságú tranzakciók is létrejöttek. Meglepő eredmény, hogy mind az érték, mind a vá-

rakozási idő korrelálatlan a távolsággal. A szórásanalízis segítségével pedig azt kaptuk, hogy a kontinenseken belüli tranzakcióknál az érték esetén nincsen, míg a távolságnál van szignifikáns különbség. A fejezet végén a sok tranzakciót lebonyolító felhasználók deanonimizálására került sor. Sikerült felderíteni a rendszer több fontos szereplőjét, ezáltal képet kapni a földrajzi elhelyezkedésükről.

3.6. Kitekintés

További vizsgálat tárgyát képezheti, ha a blokklánc letöltésével a bitcoinok mozgását nem csak kettő, hanem hosszabb láncban is elemezni tudjuk. Ekkor a készpénzhez hasonlóan lehetne egy eloszlást meghatározni a várakozási idő és a távolság segítségével.

A Bitcoin felhasználók földrajzi helyzetének eloszlása sokat változott 2013 óta, ugyanis több fejlődő országban is egyre népszerűbbé vált a Bitcoin. Ezt újra megvizsgálva valószínűleg teljesen más eredmény jönne ki a mostanihoz képest.

Még több cím deanonimizálása jobb rálátást nyújtana arra, hogy kik között mennek a tranzakciók, és hogyan alakulhatnak ki a nagy távolságok. Mivel a 2013-ban működő szolgáltatók jelentős része megszűnt, ezért ez nem egy egyszerű feladat, ugyanis csak a korábbi adatokat lehet felhasználni, mert nem bonyolítottak le új tranzakciókat.

Végül pedig érdemes lehet megvizsgálni a többi kriptovaluta hálózatát, például az Ethereumét, és összehasonlítani a Bitcoinéval. Mivel működés és felhasználás szempontjából is nagyon változatos kriptovaluták jelentek meg, ezért mindenképpen érdekes lehet ez az összevetés.

Irodalomjegyzék

[1] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *Nature* **401**, 130–131 (1999). https://www.nature.com/articles/43601, 1999.

[2] Andreas M. Antonopoulos. Mastering bitcoin. https://www.oreilly.com/library/view/mastering-bitcoin/9781491902639/ch07.html, 2014.

[3] Brockmann, Hufnagel, and Geisel. The scaling laws of human travel. *Nature* **439**, 462–465 (2006). https://www.nature.com/articles/nature04292, 2006.

[4] Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *Nat Commun* **10**, 1017 (2019). https://www.nature.com/articles/s41467-019-08746-5, 2019.

[5] Kevin Chu. Anomalous (sub) diffusion: Scaling laws. https://ocw.mit.edu/courses/18-366-random-walks-and-diffusion-fall-2006/edf52635b232991d4f943df29b8fdcdd_lecture17.pdf, 2003.

[6] John Edwards. Bitcoin's price history. https://www.investopedia.com/articles/forex/121815/bitcoins-price-history.asp.

[7] Juhász, Stéger, Kondor, and Vattay. A bayesian approach to identify bitcoin users. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0207000, 2018.

[8] Yusheng Li, Yilun Shang, and Yiting Yang. Clustering coefficients of large networks. https://researchportal.northumbria.ac.uk/ws/portalfiles/portal/17418570/Clustering_coefficients_2nd_revised.pdf, 2017.

[9] Meiklejohn, Pomarole, Jordan, Levchenko, McCoy, Voelker, and Savage. A fistful of bitcoins: Characterizing payments among men with no names. https://cseweb.ucsd.edu/~smeiklejohn/files/imc13.pdf, 2013.

[10] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf, 2008.

[11] Narayanan, Bonneau, Felten, Miller, and Goldfeder. Bitcoin and cryptocurrency technologies. https://www.lopp.net/pdf/princeton_bitcoin_book.pdf, 2016.

[12] Prakhar. Haversine formula to find distance between two points on a sphere. https://www.geeksforgeeks.org/haversine-formula-to-find-distance-between-two-points-on-a-sphere/.

[13] Wikipedia. Haversine formula. https://en.wikipedia.org/wiki/Haversine_formula.

[14] Backhausz Ágnes. Szórásanalízis és idősorok bevezetés. https://backhauszagi.web.elte.hu/gyak/sst_st4ea_k/sst_st4je_k11.pdf.